

# Recombinational Cloning Using Nucleic Acids Having Recombination Sites

## CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of the filing date of U.S. Provisional Application No. 60/065,930, filed October 24, 1997, the disclosure of which is incorporated by reference herein in its entirety. The present application is also related to U.S. Application No. 08/663,002, filed June 7, 1996, and to U.S. Application No. 08/486,139, filed June 7, 1995, the disclosures of which applications are incorporated by reference herein in their entireties.

## BACKGROUND OF THE INVENTION

### *Field of the Invention*

The present invention relates to recombinant DNA technology. DNA and vectors having engineered recombination sites are provided for use in a recombinational cloning method that enables efficient and specific recombination of DNA segments using recombination proteins. The DNAs, vectors and methods are useful for a variety of DNA exchanges, such as subcloning of DNA, *in vitro* or *in vivo*.

### *Related Art*

***Site-specific recombinases.*** Site-specific recombinases are proteins that are present in many organisms (e.g. viruses and bacteria) and have been characterized to have both endonuclease and ligase properties. These recombinases (along with associated proteins in some cases) recognize specific sequences of bases in DNA and exchange the DNA segments flanking those segments. The recombinases and associated proteins are collectively referred to as "recombination proteins" (see, e.g., Landy, A., *Current Opinion in Biotechnology* 3:699-707 (1993)).

Numerous recombination systems from various organisms have been described. See, e.g., Hoess *et al.*, *Nucleic Acids Research* 14(6):2287 (1986); Abremski *et al.*, *J. Biol. Chem.* 261(1):391 (1986); Campbell, *J. Bacteriol.* 174(23):7495 (1992); Qian *et al.*, *J. Biol. Chem.* 267(11):7794 (1992); Araki *et al.*, *J. Mol. Biol.* 225(1):25 (1992); Maeser and Kahnmann *Mol. Gen. Genet.* 230:170-176 (1991); Esposito *et al.*, *Nucl. Acids Res.* 25(18):3605 (1997).

Many of these belong to the integrase family of recombinases (Argos *et al.* *EMBO J.* 5:433-440 (1986)). Perhaps the best studied of these are the Integrase/*att* system from bacteriophage  $\lambda$  (Landy, A. *Current Opinions in Genetics and Devel.* 3:699-707 (1993)), the Cre/*loxP* system from bacteriophage P1 (Hoess and Abremski (1990) In *Nucleic Acids and Molecular Biology*, vol. 4. Eds.: Eckstein and Lilley, Berlin-Heidelberg: Springer-Verlag; pp. 90-109), and the FLP/FRT system from the *Saccharomyces cerevisiae* 2  $\mu$  circle plasmid (Broach *et al.* *Cell* 29:227-234 (1982)).

While these recombination systems have been characterized for particular organisms, the related art has only taught using recombinant DNA flanked by recombination sites, for *in vivo* recombination.

Backman (U.S. Patent No. 4,673,640) discloses the *in vivo* use of  $\lambda$  recombinase to recombine a protein producing DNA segment by enzymatic site-specific recombination using wild-type recombination sites *attB* and *attP*.

Hasan and Szybalski (*Gene* 56:145-151 (1987)) discloses the use of  $\lambda$  Int recombinase *in vivo* for intramolecular recombination between wild type *attP* and *attB* sites which flank a promoter. Because the orientations of these sites are inverted relative to each other, this causes an irreversible flipping of the promoter region relative to the gene of interest.

Palazzolo *et al.* *Gene* 88:25-36 (1990), discloses phage lambda vectors having bacteriophage  $\lambda$  arms that contain restriction sites positioned outside a cloned DNA sequence and between wild-type *loxP* sites. Infection of *E. coli* cells that express the Cre recombinase with these phage vectors results in recombination between the *loxP* sites and the *in vivo* excision of the plasmid replicon, including the cloned cDNA.

Pósfai *et al.* (*Nucl. Acids Res.* 22:2392-2398 (1994)) discloses a method for inserting into genomic DNA partial expression vectors having a selectable marker, flanked by two wild-type FRT recognition sequences. FLP site-specific recombinase as present in the cells is used to integrate the vectors into the genome at predetermined sites. Under conditions where the replicon is functional, this cloned genomic DNA can be amplified.

Bebee *et al.* (U.S. Patent No. 5,434,066) discloses the use of site-specific recombinases such as Cre for DNA containing two *loxP* sites is used for *in vivo* recombination between the sites.

Boyd (*Nucl. Acids Res.* 21:817-821 (1993)) discloses a method to facilitate the cloning of blunt-ended DNA using conditions that encourage intermolecular ligation to a dephosphorylated vector that contains a wild-type *loxP* site acted upon by a Cre site-specific recombinase present in *E. coli* host cells.

Waterhouse *et al.* (PCT No. 93/19172 and *Nucleic Acids Res.* 21 (9):2265 (1993)) disclose an *in vivo* method where light and heavy chains of a particular antibody were cloned in different phage vectors between *loxP* and *loxP 511* sites and used to transfect new *E. coli* cells. Cre, acting in the host cells on the two parental molecules (one plasmid, one phage), produced four products in equilibrium: two different cointegrates (produced by recombination at either *loxP* or *loxP 511* sites), and two daughter molecules, one of which was the desired product.

In contrast to the other related art, Schlake & Bode (*Biochemistry* 33:12746-12751 (1994)) discloses an *in vivo* method to exchange expression cassettes at defined chromosomal locations, each flanked by a wild type and a spacer-mutated FRT recombination site. A double-reciprocal crossover was mediated in cultured mammalian cells by using this FLP/FRT system for site-specific recombination.

**Transposases.** The family of enzymes, the transposases, has also been used to transfer genetic information between replicons. Transposons are structurally variable, being described as simple or compound, but typically encode the recombinase gene flanked by DNA sequences organized in inverted orientations. Integration of transposons can be random or highly specific.

Representatives such as Tn7, which are highly site-specific, have been applied to the *in vivo* movement of DNA segments between replicons (Lucklow *et al.*, *J. Virol.* 67:4566-4579 (1993)).

Devine and Boeke *Nucl. Acids Res.* 22:3765-3772 (1994), discloses the construction of artificial transposons for the insertion of DNA segments, *in vitro*, into recipient DNA molecules. The system makes use of the integrase of yeast TY1 virus-like particles. The DNA segment of interest is cloned, using standard methods, between the ends of the transposon-like element TY1. In the presence of the TY1 integrase, the resulting element integrates randomly into a second target DNA molecule.

**DNA cloning.** The cloning of DNA segments currently occurs as a daily routine in many research labs and as a prerequisite step in many genetic analyses. The purpose of these clonings is various, however, two general purposes can be considered: (1) the initial cloning of DNA from large DNA or RNA segments (chromosomes, YACs, PCR fragments, mRNA, etc.), done in a relative handful of known vectors such as pUC, pGem, pBlueScript, and (2) the subcloning of these DNA segments into specialized vectors for functional analysis. A great deal of time and effort is expended both in the transfer of DNA segments from the initial cloning vectors to the more specialized vectors. This transfer is called subcloning.

The basic methods for cloning have been known for many years and have changed little during that time. A typical cloning protocol is as follows:

- (1) digest the DNA of interest with one or two restriction enzymes;
- (2) gel purify the DNA segment of interest when known;
- (3) prepare the vector by cutting with appropriate restriction enzymes, treating with alkaline phosphatase, gel purify etc., as appropriate;
- (4) ligate the DNA segment to the vector, with appropriate controls to eliminate background of uncut and self-ligated vector;
- (5) introduce the resulting vector into an *E. coli* host cell;
- (6) pick selected colonies and grow small cultures overnight;
- (7) make DNA minipreps; and

(8) analyze the isolated plasmid on agarose gels (often after diagnostic restriction enzyme digestions) or by PCR.

The specialized vectors used for subcloning DNA segments are functionally diverse. These include but are not limited to: vectors for expressing genes in various organisms; for regulating gene expression; for providing tags to aid in protein purification or to allow tracking of proteins in cells; for modifying the cloned DNA segment (*e.g.*, generating deletions); for the synthesis of probes (*e.g.*, riboprobes); for the preparation of templates for DNA sequencing; for the identification of protein coding regions; for the fusion of various protein-coding regions; to provide large amounts of the DNA of interest, *etc.* It is common that a particular investigation will involve subcloning the DNA segment of interest into several different specialized vectors.

As known in the art, simple subclonings can be done in one day (*e.g.*, the DNA segment is not large and the restriction sites are compatible with those of the subcloning vector). However, many other subclonings can take several weeks, especially those involving unknown sequences, long fragments, toxic genes, unsuitable placement of restriction sites, high backgrounds, impure enzymes, *etc.* Subcloning DNA fragments is thus often viewed as a chore to be done as few times as possible. Several methods for facilitating the cloning of DNA segments have been described, *e.g.*, as in the following references.

Ferguson, J., *et al. Gene* 16:191 (1981), discloses a family of vectors for subcloning fragments of yeast DNA. The vectors encode kanamycin resistance. Clones of longer yeast DNA segments can be partially digested and ligated into the subcloning vectors. If the original cloning vector conveys resistance to ampicillin, no purification is necessary prior to transformation, since the selection will be for kanamycin.

Hashimoto-Gotoh, T., *et al. Gene* 41:125 (1986), discloses a subcloning vector with unique cloning sites within a streptomycin sensitivity gene; in a streptomycin-resistant host, only plasmids with inserts or deletions in the dominant sensitivity gene will survive streptomycin selection.

Accordingly, traditional subcloning methods, using restriction enzymes and ligase, are time consuming and relatively unreliable. Considerable labor is

expended, and if two or more days later the desired subclone can not be found among the candidate plasmids, the entire process must then be repeated with alternative conditions attempted. Although site specific recombinases have been used to recombine DNA *in vivo*, the successful use of such enzymes *in vitro* was expected to suffer from several problems. For example, the site specificities and efficiencies were expected to differ *in vitro*; topologically-linked products were expected; and the topology of the DNA substrates and recombination proteins was expected to differ significantly *in vitro* (see, e.g., Adams *et al*, *J. Mol. Biol.* 226:661-73 (1992)). Reactions that could go on for many hours *in vivo* were expected to occur in significantly less time *in vitro* before the enzymes became inactive. Multiple DNA recombination products were expected in the biological host used, resulting in unsatisfactory reliability, specificity or efficiency of subcloning. Thus, *in vitro* recombination reactions were not expected to be sufficiently efficient to yield the desired levels of product.

Accordingly, there is a long felt need to provide an alternative subcloning system that provides advantages over the known use of restriction enzymes and ligases.

#### SUMMARY OF THE INVENTION

The present invention provides nucleic acids, vectors and methods for obtaining amplified, chimeric or recombinant nucleic acid molecules using recombination proteins and at least one recombination site, *in vitro* or *in vivo*. These methods are highly specific, rapid, and less labor intensive than standard cloning or subcloning techniques. The improved specificity, speed and yields of the present invention facilitates DNA or RNA cloning or subcloning, regulation or exchange useful for any related purpose.

The present invention relates to nucleic acids, vectors and methods for moving or exchanging nucleic acid segments (preferably DNA segments or fragments) using at least one recombination site and at least one recombination protein to provide chimeric DNA molecules which have the desired characteristic(s) and/or DNA segment(s). Use of the invention thus allows for

cloning or subcloning such nucleic acid molecules into a variety of vectors. Generally, one or more parent nucleic acid molecules (preferably DNA molecules) are recombined to give one or more daughter molecules, at least one of which is the desired Product molecule, which is preferably a vector comprising the desired nucleic acid segment. The invention thus relates to nucleic acid molecules, vectors and methods to effect the exchange and/or to select for one or more desired products.

One embodiment of the present invention relates to a method of making chimeric molecule, which comprises

(a) combining *in vitro* or *in vivo*

(i) one or more Insert Donor molecules comprising a desired nucleic acid segment flanked by a first recombination site and a second recombination site, wherein the first and second recombination sites do not substantially recombine with each other;

(ii) one or more Vector Donor molecules comprising a third recombination site and a fourth recombination site, wherein the third and fourth recombination sites do not substantially recombine with each other; and

(iii) one or more site specific recombination proteins capable of recombining the first and third recombinational sites and/or the second and fourth recombinational sites;

thereby allowing recombination to occur, so as to produce at least one cointegrate nucleic acid molecule, at least one desired Product nucleic acid molecule which comprises said desired segment, and optionally a Byproduct nucleic acid molecule; and then, optionally,

(b) selecting for the Product or Byproduct DNA molecule.

In another embodiment, the present invention relates to a method of making chimeric molecule, which comprises

(a) combining *in vitro* or *in vivo*

(i) one or more Insert Donor molecules comprising a desired nucleic acid segment flanked by two or more recombination sites wherein said recombination sites do not substantially recombine with each other;

(ii) one or more Vector Donor molecules comprising two or more recombination sites, wherein said recombination sites do not substantially recombine with each other; and

(iii) one or more site specific recombination proteins;

5 (b) incubating said combination under conditions sufficient to transfer one or more said desired segments into one or more of said Vector Donor molecules, thereby producing one or more Product molecules. The resulting Product molecules may optionally be selected or isolated away from other molecules such as cointegrate molecules, Byproduct molecules, and unreacted  
10 Vector Donor molecules or Insert Donor molecules. In a preferred aspect of the invention, the Insert Donor molecules are combined with one or more different Vector Donor molecules, thereby allowing for the production of different Product molecules in which the nucleic acid of interest is transferred into any number of different vectors in the single step.

15 In accordance with the invention, the above methods may be reversed to provide the original Insert Donor molecules which may then be used in combination with one or more different Vector Donor molecules to produce new Product or Byproduct molecules. Alternatively, the Product molecules produced by the method of the invention may serve as the Insert Donor molecules which  
20 may be used directly in combination with one or more different Vector Donor molecules, thereby producing new Product or Byproduct molecules. Thus, nucleic acid molecules of interest may be transferred or moved to any number of desired vectors, thereby providing an efficient means for subcloning molecules of interest.

25 Thus, the invention relates to combining a Product molecule with a second Vector Donor molecules to produce a second Product molecule. The second Product DNA molecule may then be utilized in combination with a third Vector Donor molecule to produce a third Product molecule. This process of the invention may be repeated any number of times to transfer or move the insert of  
30 interest into any number of different vectors. In this aspect of the invention, a combination of two or more different Vector Donor molecules may be combined with the Product molecule to produce in a single step different Product molecules



in which the desired nucleic acid segment (derived from the Product DNA molecule) is transferred into any number of different vectors.

In particular, the present invention relates to a method for cloning or subcloning one or more desired nucleic acid molecules comprising

(a) combining *in vitro* or *in vivo*

(i) one or more Insert Donor molecules comprising one or more desired nucleic acid segments flanked by at least two recombination sites, wherein said recombination sites do not substantially recombined with each other;

(ii) one or more Vector Donor molecules comprising at least two recombination sites, wherein said recombination sites do not substantially recombine with each other; and

(iii) one or more site specific recombination proteins;

(b) incubating said combination under conditions sufficient to allow one or more of said desired segments to be transferred into one or more of said Vector Donor molecules, thereby producing one or more Product molecules;

(c) optionally selecting for or isolating said Product molecule;

(d) combining *in vitro* or *in vivo*

(i) one or more of said Product molecules comprising said desired segments flanked by two or more recombination sites, wherein said recombination sites do not substantially recombine with each other;

(ii) one or more different Vector Donor molecules comprising two or more recombination sites, wherein said recombination sites do not substantially recombine with each other; and

(iii) one or more site specific recombination protein; and

(e) incubating said combination under conditions sufficient to transfer one or more of said desired segments into one or more of said different Vector Donor molecules, thereby producing one or more different Product molecules.

In accordance with the invention, Vector Donor molecules may comprise vectors which may function in a variety of systems or host cells. Preferred vectors for use in the invention include prokaryotic vectors, eukaryotic vectors or vectors which may shuttle between various prokaryotic and/or eukaryotic systems (e.g. shuttle vectors). Preferred prokaryotic vectors for use in the invention include but

are not limited to vectors which may propagate and/or replicate in gram negative and/or gram positive bacteria, including bacteria of the genus *Escherichia*, *Salmonella*, *Proteus*, *Clostridium*, *Klebsiella*, *Bacillus*, *Streptomyces*, and *Pseudomonas* and preferably in the species *E. coli*. Eukaryotic vectors for use in the invention include vectors which propagate and/or replicate and yeast cells, plant cells, mammalian cells, (particularly human), fungal cells, insect cells, fish cells and the like. Particular vectors of interest include but are not limited to cloning vectors, sequencing vectors, expression vectors, fusion vectors, two-hybrid vectors, gene therapy vectors, and reverse two-hybrid vectors. Such vectors may be used in prokaryotic and/or eukaryotic systems depending on the particular vector.

The Insert Donor molecules used in accordance with the invention preferably comprise two or more recombination sites which allow the insert (e.g. the nucleic acid segment of interest) of the Donor molecules to be transferred or moved into one or more Vector Donor molecules in accordance with the invention. The Insert Donor molecules of the invention may be prepared by any number of techniques by which two or more recombination sites are added to the molecule of interest. Such means for including recombination sites to prepare the Insert Donor molecules of the invention includes mutation of a nucleic acid molecule (e.g. random or site specific mutagenesis), recombinant techniques (e.g. ligation of adapters or nucleic acid molecules comprising recombination sites to linear molecules), amplification (e.g. using primers which comprise recombination sites or portions thereof) transposition (e.g. using transposons which comprise recombination sites), recombination (e.g. using one or more homologous sequences comprising recombination sites), nucleic acid synthesis (e.g. chemical synthesis of molecules comprising recombination sites or enzymatic synthesis using various polymerases or reverse transcriptases) and the like. In accordance with the invention, nucleic acid molecules to which one or more recombination sites are added may be any nucleic acid molecule derived from any source and may include non naturally occurring nucleic acids (e.g. RNA's; see U.S. Patent Nos. 5,539,082 and 5,482,836). Particularly preferred nucleic acid molecules are DNA molecules (single stranded or double stranded). Additionally, the nucleic acid

molecules of interest for producing Insert Donor molecules may be linear or circular and further may comprise a particular sequence of interest (e.g. a gene) or may be a population of molecules (e.g. molecules generated from a genomic or cDNA libraries).

Thus, the invention relates to a number of methods for preparing Insert Donor molecules and the Insert Donor molecules produced by such methods. In one aspect of the invention, primers comprising one or more recombination sites or portions thereof are used in the nucleic acid synthesis or nucleic acid amplification to prepare the Insert Donor molecules of the invention. Thus, the invention relates to a method of synthesizing a nucleic acid molecule comprising:

(a) mixing one or more nucleic acid templates with a polypeptide having polymerase activity and one or more primers comprising one or more recombination sites or portions thereof; and

(b) incubating said mixture under conditions sufficient to synthesize one or more nucleic acid molecules which are complementary to all or a portion of said templates and which comprises one or more recombination sites. In accordance with the invention, the synthesized nucleic acid molecule comprising one or more recombination sites may be used as templates under appropriate conditions to synthesize nucleic acid molecules complementary to all or a portion of the recombination site containing templates, thereby forming double stranded molecules comprising one or more recombination sites. Preferably, such second synthesis step is performed in the presence of one or more primers comprising one or more recombination sites. In yet another aspect, the synthesized double stranded molecules may be amplified using primers which may comprise one or more recombination sites.

In another aspect of the invention, one or more recombination sites may be added to nucleic acid molecules by any of a number of nucleic acid amplification techniques. In particular, such method comprises:

(a) contacting a first nucleic acid molecule with a first primer molecule which is complementary to a portion of said first nucleic acid molecule and a second nucleic acid molecule with a second primer molecule which is

complementary to a portion of said second nucleic acid molecule in the presence of one or more polypeptides having polymerases activity;

(b) incubating said molecules under conditions sufficient to form a third nucleic acid molecule complementary to all or a portion of said first nucleic acid molecule and the fourth nucleic acid molecule complementary to all or a portion of said second nucleic acid molecule;

(c) denaturing said first and third and said second and fourth nucleic acid molecules; and

(d) repeating steps (a) through (c) one or more times, wherein said first and/or said second primer molecules comprise one or more recombination sites or portions thereof.

In yet another aspect of the invention, a method for adding one or more recombination sites to nucleic acid molecules may comprise:

(a) contacting one or more nucleic acid molecules with one or more adapters or nucleic acid molecules which comprise one or more recombination sites or portions thereof; and

(b) incubating said mixture under conditions sufficient to add one or more recombination sites to said nucleic acid molecules. Preferably, linear molecules are used for adding such adapters or molecules in accordance with the invention and such adapters or molecules are preferably added to one or more termini of such linear molecules. The linear molecules may be prepared by any technique including mechanical (e.g. sonication or shearing) or enzymatic (e.g. nucleases such as restriction endonucleases). Thus, the method of the invention may further comprise digesting the nucleic acid molecule with one or more nucleases (preferably any restriction endonucleases) and ligating one or more of the recombination site containing adapters or molecules to the molecule of interest. Ligation may be accomplished using blunt ended or stick ended molecules. Alternatively, topoisomerases may be used to introduce recombination sites in accordance with the invention. Topoisomerases cleave and rejoin nucleic acid molecules and therefore may be used in place of nucleases and ligases.

In another aspect, one or more recombination sites may be added to nucleic acid molecules by *de novo* synthesis. Thus, the invention relates to such

a method which comprises chemically synthesizing one or more nucleic acid molecules in which recombination sites are added by adding the appropriate sequence of nucleotides during the synthesis process.

In another embodiment of the invention, one or more recombination sites may be added to nucleic acid molecules of interest by a method which comprises:

(a) contacting one or more nucleic acid molecules with one or more integration sequences which comprise one or more recombination sites or portions thereof; and

(b) incubation of said mixture under conditions sufficient to incorporate said recombination site containing integration sequences into said nucleic acid molecules. In accordance with this aspect of the invention, integration sequences may comprise any nucleic acid molecules which through recombination or by integration become a part of the nucleic acid molecule of interest. Integration sequences may be introduced in accordance with this aspect of the invention by *in vivo* or *in vitro* recombination (homologous recombination or illegitimate recombination) or by *in vivo* or *in vitro* installation by using transposons, insertion sequences, integrating viruses, homing introns, or other integrating elements.

In another aspect, the invention relates to kits for carrying out the methods of the invention and more specifically relates to cloning or subcloning kits and kits for making Insert Donor molecules of the invention. Such kits may comprise a carrier or receptacle being compartmentalized to receive and hold therein any number of containers. Such containers may contain any number of components for carrying out the methods of the invention or combinations of such components. In particular, a kit of the invention may comprise one or more components (or combinations thereof) selected from the group consisting of one or more recombination proteins or recombinases, one or more Vector Donor molecules, one or more Insert Donor molecules and one or more host cells (e.g. competent cells).

Kits for making the Insert Donor molecules of the invention may comprise any or a number of components and the composition of such kits may vary depending on the specific method involved. Kits for synthesizing Insert Donor

5 molecules by amplification may comprise one or more components (or combinations thereof) selected from the group consisting of one or more polypeptides having polymerase activity (preferably DNA polymerases and most preferably thermostable DNA polymerases), one or more nucleotides, and one or more primers comprising one or more recombination sites. Kits for inserting or adding recombination sites to nucleic acid molecules of interest may comprise one or more nucleases (preferably restriction endonucleases), one or more ligases, one or more topoisomerases one or more polymerases, and one or more nucleic acid molecules or adapters comprising one or more recombination sites. Kits for integrating recombination sites into one or more nucleic acid molecules of interest may comprise one or more components (or combinations thereof) selected from the group consisting of one or more integration sequences comprising one or more recombination sites. Such integration sequences may comprise one or more transposons, integrating viruses, homologous recombination sequences, one or more host cells and the like.

10 The invention also relates to compositions for carrying out the methods of the invention or compositions which are produced from carrying out the methods of the invention. In particular, such compositions may comprise one or more Insert Donor molecules, one or more Vector Donor molecules and one or more recombination proteins (or combinations thereof). In a further aspect, the compositions of the invention may comprise one or more cointegrate molecules, one or more Product molecules and one or more Byproduct molecule (or combinations thereof).

20 Compositions related to preparing Insert Donor molecules may vary depending on the particular method utilized in preparing the desired Insert Donor molecules. Compositions for preparing such molecules by amplification may comprise one or more polypeptides having polymerase activity, one or more primers comprising one or more recombination sites, one or more nucleotides and one or more nucleic acid molecule to be amplified (or combinations thereof). Compositions related to inserting or adding recombination sites in a desired nucleic acid molecule may comprise one or more nucleic acid molecules or adapters comprising one or more recombination sites, one or more ligases, one or

more restriction endonucleases, one or more topoisomerases, and one or more nucleic acid molecules desired to contain such recombination sites (or combinations thereof). Compositions related to integration of recombination sites in a desired nucleic acid molecule may comprise one or more integration sequences comprising one or more recombination sites and one or more nucleic acid molecules desired to contain the recombination sites.

In a particularly preferred aspect of the invention, libraries (e.g. populations of genomic DNA or cDNA, or populations of nucleic acid molecules, produced by *de novo* synthesis such as random sequences or degenerate oligonucleotides) are utilized in accordance with the present invention. By inserting or adding recombination sites to such populations of nucleic acid molecules, a population of Insert Donor molecules are produced. By the recombination methods of the invention, the library may be easily moved into different vectors (or combinations of vectors) and thus into different host systems (prokaryotic and eukaryotic) to evaluate and analyze the library or a particular sequences or clones derived from the library. Alternatively, the vectors containing the desired molecule may be used *in vitro* systems such as *in vitro* expression systems for production of RNA and/or protein. In a particularly preferred aspect, one or more recombination sites are added to nucleic acid molecules of the library by method comprising:

(a) mixing a population of linear nucleic acid molecules with one or more adapters comprising one or more recombination sites; and

(b) incubating said mixture under conditions sufficient to add one or more of said adapters to one or more termini of said linear molecules. In a preferred aspect, the population of nucleic acid molecules are double stranded DNA molecules (preferably genomic DNA or cDNA). A population of linear fragments for use in the invention may be prepared by cleaving (by mechanical or enzymatic means) the genomic or cDNA. In a preferred aspect, the adapters are added to one or more termini of the linear molecules.

In another particularly preferred aspect of the invention, cDNA libraries are used to prepare a population of Insert Donor DNA molecules of the invention. In particular, this aspect of the invention relates to a method which comprises:

(a) contacting a population of RNA, mRNA or polyA+ RNA templates with one or more polypeptides having reverse transcriptase activity and one or more primers which comprises one or more recombination sites;

5 (b) incubating said mixture under conditions sufficient to synthesize a first population of DNA molecules complementary to said templates, wherein said DNA molecules comprise one or more recombination sites. This aspect of the invention may further comprise incubating said synthesized DNA under conditions sufficient to make a second population of DNA molecules complementary to all or a portion of said first population of DNA molecules, thereby forming a population of double stranded DNA molecules comprising one or more recombination sites.

10 In a particularly preferred aspect, the Insert Donor molecules of the invention comprise at least two recombination sites and where the Insert Donor molecules are linear, such two or more recombination sites are preferably located at or near both termini of the molecules. In accordance with the invention, the use of additional recombination sites (i.e. more than two) may be used to facilitate subcloning of different inserts within the Insert Donor molecule, depending on the type and placement of such recombination sites.

15 Other embodiments include DNA and vectors useful in the methods of the present invention. In particular, Vector Donor molecules are provided in one embodiment, wherein DNA segments within the Vector Donor are separated either by, (i) in a circular Vector Donor, at least two recombination sites, or (ii) in a linear Vector Donor, at least one recombination site, where the recombination sites are preferably engineered to enhance specificity or efficiency of recombination. One Vector Donor embodiment comprises a first DNA segment and a second DNA segment, the first or second segment comprising a selectable marker. A second Vector Donor embodiment comprises a first DNA segment and a second DNA segment, the first or second DNA segment comprising a toxic gene. A third Vector Donor embodiment comprises a first DNA segment and a second DNA segment, the first or second DNA segment comprising an inactive fragment of at least one selectable marker, wherein the inactive fragment of the Selectable marker is capable of reconstituting a functional

20

25

30



Selectable marker when recombined across the first or second recombination site with another inactive fragment of at least one Selectable marker.

Other preferred embodiments of the present invention will be apparent to one of ordinary skill in light of what is known in the art, in light of the following drawings and description of the invention, and in light of the claims.

## BRIEF DESCRIPTION OF THE DRAWINGS

**Figure 1** depicts one general method of the present invention, wherein the starting (parent) DNA molecules can be circular or linear. The goal is to exchange the new subcloning vector D for the original cloning vector B. It is desirable in one embodiment to select for AD and against all the other molecules, including the Cointegrate. The square and circle are sites of recombination: *e.g.*, *loxP* sites, *att* sites, *etc.* For example, segment D can contain expression signals, new drug markers, new origins of replication, or specialized functions for mapping or sequencing DNA.

**Figure 2A** depicts an *in vitro* method of recombining an Insert Donor plasmid (here, pEZC705) with a Vector Donor plasmid (here, pEZC726), and obtaining Product DNA and Byproduct daughter molecules. The two recombination sites are *attP* and *loxP* on the Vector Donor. On one segment defined by these sites is a kanamycin resistance gene whose promoter has been replaced by the *tetOP* operator/promoter from transposon *Tn10*. *See, e.g.*, Sizemore *et al.*, *Nucl. Acids Res.* 18(10):2875 (1990). In the absence of tet repressor protein, *E. coli* RNA polymerase transcribes the kanamycin resistance gene from the *tetOP*. If tet repressor is present, it binds to *tetOP* and blocks transcription of the kanamycin resistance gene. The other segment of pEZC726 has the *tet* repressor gene expressed by a constitutive promoter. Thus cells transformed by pEZC726 are resistant to chloramphenicol, because of the chloramphenicol acetyl transferase gene on the same segment as *tetR*, but are sensitive to kanamycin. The recombinase-mediated reactions result in separation of the *tetR* gene from the regulated kanamycin resistance gene. This separation results in kanamycin resistance only in cells receiving the desired recombination

product. The first recombination reaction is driven by the addition of the recombinase called Integrase. The second recombination reaction is driven by adding the recombinase Cre to the Cointegrate (here, pEYC7 Cointegr).

**Figure 2B** depicts a restriction map of pEYC705.

**Figure 2C** depicts a restriction map of pEYC726.

**Figure 2D** depicts a restriction map of pEYC7 Coint.

**Figure 2E** depicts a restriction map of Intprod.

**Figure 2F** depicts a restriction map of Intbypro.

**Figure 3A** depicts an *in vitro* method of recombining an Insert Donor plasmid (here, pEYC602) with a Vector Donor plasmid (here, pEYC629), and obtaining Product (here, EYC6prod) and Byproduct (here, EYC6Bypr) daughter molecules. The two recombination sites are *loxP* and *loxP 511*. One segment of pEYC629 defined by these sites is a kanamycin resistance gene whose promoter has been replaced by the *tetOP* operator/promoter from transposon *Tn10*. In the absence of tet repressor protein, *E. coli* RNA polymerase transcribes the kanamycin resistance gene from the *tetOP*. If tet repressor is present, it binds to *tetOP* and blocks transcription of the kanamycin resistance gene. The other segment of pEYC629 has the *tet* repressor gene expressed by a constitutive promoter. Thus cells transformed by pEYC629 are resistant to chloramphenicol, because of the chloramphenicol acetyl transferase gene on the same segment as *tetR*, but are sensitive to kanamycin. The reactions result in separation of the *tetR* gene from the regulated kanamycin resistance gene. This separation results in kanamycin resistance in cells receiving the desired recombination product. The first and the second recombination events are driven by the addition of the same recombinase, Cre.

**Figure 3B** depicts a restriction map of EYC6Bypr.

**Figure 3C** depicts a restriction map of EYC6prod.

**Figure 3D** depicts a restriction map of pEYC602.

**Figure 3E** depicts a restriction map of pEYC629.

**Figure 3F** depicts a restriction map of EYC6coint.

**Figure 4A** depicts an application of the *in vitro* method of recombinational cloning to subclone the chloramphenicol acetyl transferase gene into a vector for expression in eukaryotic cells. The Insert Donor plasmid, pEYC843, is comprised of the chloramphenicol acetyl transferase gene of *E. coli*, cloned between *loxP* and *attB* sites such that the *loxP* site is positioned at the 5'-end of the gene. The Vector Donor plasmid, pEYC1003, contains the cytomegalovirus eukaryotic promoter apposed to a *loxP* site. The supercoiled plasmids were combined with lambda Integrase and Cre recombinase *in vitro*. After incubation, competent *E. coli* cells were transformed with the recombinational reaction solution. Aliquots of transformations were spread on agar plates containing kanamycin to select for the Product molecule (here CMVProd).

**Figure 4B** depicts a restriction map of pEYC843.

**Figure 4C** depicts a restriction map of pEYC1003.

**Figure 4D** depicts a restriction map of CMVBypro.

**Figure 4E** depicts a restriction map of CMVProd.

**Figure 4F** depicts a restriction map of CMVcoint.

**Figure 5A** depicts a vector diagram of pEYC1301.

**Figure 5B** depicts a vector diagram of pEYC1305.

**Figure 5C** depicts a vector diagram of pEYC1309.

**Figure 5D** depicts a vector diagram of pEYC1313.

**Figure 5E** depicts a vector diagram of pEYC1317.

**Figure 5F** depicts a vector diagram of pEYC1321.

**Figure 5G** depicts a vector diagram of pEYC1405.

**Figure 5H** depicts a vector diagram of pEYC1502.

**Figure 6A** depicts a vector diagram of pEYC1603.

**Figure 6B** depicts a vector diagram of pEYC1706.

**Figure 7A** depicts a vector diagram of pEYC2901.

**Figure 7B** depicts a vector diagram of pEYC2913.

**Figure 7C** depicts a vector diagram of pEYC3101.

**Figure 7D** depicts a vector diagram of pEYC1802.

**Figure 8A** depicts a vector diagram of pGEX-2TK.

**Figure 8B** depicts a vector diagram of pEZC3501.

**Figure 8C** depicts a vector diagram of pEZC3601.

**Figure 8D** depicts a vector diagram of pEZC3609.

**Figure 8E** depicts a vector diagram of pEZC3617.

**Figure 8F** depicts a vector diagram of pEZC3606.

**Figure 8G** depicts a vector diagram of pEZC3613.

**Figure 8H** depicts a vector diagram of pEZC3621.

**Figure 8I** depicts a vector diagram of GST-CAT.

**Figure 8J** depicts a vector diagram of GST-phoA.

**Figure 8K** depicts a vector diagram of pEZC3201.

**Figure 9A** depicts a diagram of 5.2kb PCR prod.

**Figure 9B** depicts a vector diagram of pEZC1202.

**Figure 9C** depicts a vector diagram of 5.2kb clone.

**Figure 10A** depicts a vector diagram of pEZC5601.

**Figure 10B** depicts a vector diagram of pEZC6701.

**Figure 10C** depicts a vector diagram of attL product.

**Figure 10D** depicts attR product.

**Figure 11A** depicts a vector diagram of pEZC7102.

**Figure 11B** depicts a vector diagram of pEZC7501.

**Figure 11C** depicts the attL product.

**Figure 12A** depicts an amp PCR product with terminal attB sites.

**Figure 12B** depicts a tet PCR product with terminal attB sites.

**Figure 12C** depicts a restriction map of amp7102.

**Figure 12D** depicts a restriction map of tet 7102.

## DETAILED DESCRIPTION OF THE INVENTION

It is unexpectedly discovered in the present invention that reversible and/or repeatable cloning and subcloning reactions can be used to manipulate nucleic acids to form chimeric nucleic acids using recombination proteins and recombination sites. Recombinational cloning according to the present invention

thus uses recombination proteins with recombinant nucleic acid molecules having at least one selected recombination site for moving or exchanging segments of nucleic acid molecules, *in vitro* and *in vivo*.

These methods use recombination reactions to generate chimeric DNA or RNA molecules that have the desired characteristic(s) and/or nucleic acid segment(s). The methods of the invention provide a means in which nucleic acid molecule of interest may be moved or transferred into any number of vector systems. In accordance with the invention, such transfer to various vector systems may be accomplished separately, sequentially or in mass (e.g. into any number of different vectors in one step). The improved specificity, speed and/or yields of the present invention facilitates DNA or RNA cloning, subcloning, regulation or exchange useful for any related purpose. Such purposes include *in vitro* recombination of DNA or RNA segments and *in vitro* or *in vivo* insertion or modification of transcribed, replicated, isolated or genomic DNA or RNA.

### Definitions

In the description that follows, a number of terms used in recombinant DNA technology are utilized extensively. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

**Byproduct:** is a daughter molecule (a new clone produced after the second recombination event during the recombinational cloning process) lacking the segment which is desired to be cloned or subcloned.

**Cointegrate:** is at least one recombination intermediate nucleic acid molecule of the present invention that contains both parental (starting) molecules. It will usually be circular. In some embodiments it can be linear.

**Host:** is any prokaryotic or eukaryotic organism that can be a recipient of the recombinational cloning Product. A "host," as the term is used herein, includes prokaryotic or eukaryotic organisms that can be genetically engineered. For examples of such hosts, see Maniatis *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1982).

**Insert or Inserts:** include the desired nucleic acid segment or a population of nucleic acid segments (segment *A* of Figure 1) which may be manipulated by the methods of the present invention. Thus, the terms Insert(s) are meant to include a particular nucleic acid (preferably DNA) segment or a population of segments. Such Insert(s) can comprise one or more genes.

**Insert Donor:** is one of the two parental nucleic acid molecules (e.g. RNA or DNA) of the present invention which carries the Insert. The Insert Donor molecule comprises the Insert flanked on both sides with recombination sites. The Insert Donor can be linear or circular. In one embodiment of the invention, the Insert Donor is a circular DNA molecule and further comprises a cloning vector sequence outside of the recombination signals (see Figure 1). When a population of Inserts or population of nucleic acid segments are used to make the Insert Donor, a population of Insert Donors result and may be used in accordance with the invention.

**Product:** is one the desired daughter molecules comprising the *A* and *D* sequences which is produced after the second recombination event during the recombinational cloning process (see Figure 1). The Product contains the nucleic acid which was to be cloned or subcloned. In accordance with the invention, when a population of Insert Donors are used, the resulting population of Product molecules will contain all or a portion of the population of Inserts of the Insert Donors and preferably will contain a representative population of the original molecules of the Insert Donors.

**Promoter:** is a DNA sequence generally described as the 5'-region of a gene, located proximal to the start codon. The transcription of an adjacent DNA segment is initiated at the promoter region. A repressible promoter's rate of transcription decreases in response to a repressing agent. An inducible promoter's rate of transcription increases in response to an inducing agent. A constitutive promoter's rate of transcription is not specifically regulated, though it can vary under the influence of general metabolic conditions.

**Recognition sequence:** Recognition sequences are particular sequences which a protein, chemical compound, DNA, or RNA molecule (*e.g.*, restriction

endonuclease, a modification methylase, or a recombinase) recognizes and binds. In the present invention, a recognition sequence will usually refer to a recombination site. For example, the recognition sequence for Cre recombinase is *loxP* which is a 34 base pair sequence comprised of two 13 base pair inverted repeats (serving as the recombinase binding sites) flanking an 8 base pair core sequence. See Figure 1 of Sauer, B., *Current Opinion in Biotechnology* 5:521-527 (1994). Other examples of recognition sequences are the *attB*, *attP*, *attL*, and *attR* sequences which are recognized by the recombinase enzyme  $\lambda$  Integrase. *attB* is an approximately 25 base pair sequence containing two 9 base pair core-type Int binding sites and a 7 base pair overlap region. *attP* is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins integration host factor (IHF), FIS and excisionase (Xis). See Landy, *Current Opinion in Biotechnology* 3:699-707 (1993). Such sites may also be engineered according to the present invention to enhance production of products in the methods of the invention. When such engineered sites lack the P1 or H1 domains to make the recombination reactions irreversible (e.g., *attR* or *attP*), such sites may be designated *attR'* or *attP'* to show that the domains of these sites have been modified in some way.

**Recombinase:** is an enzyme which catalyzes the exchange of DNA segments at specific recombination sites.

**Recombinational Cloning:** is a method described herein, whereby segments of nucleic acid molecules or populations of such molecules are exchanged, inserted, replaced, substituted or modified, *in vitro* or *in vivo*.

**Recombination proteins:** include excisive or integrative proteins, enzymes, co-factors or associated proteins that are involved in recombination reactions involving one or more recombination sites. See, Landy (1994), *infra*.

**Repression cassette:** is a nucleic acid segment that contains a repressor of a Selectable marker present in the subcloning vector.

**Selectable marker:** is a DNA segment that allows one to select for or against a molecule or a cell that contains it, often under particular conditions.

These markers can encode an activity, such as, but not limited to, production of RNA, peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds or compositions and the like. Examples of Selectable markers include but are not limited to: (1) DNA segments that encode products which provide resistance against otherwise toxic compounds (*e.g.*, antibiotics); (2) DNA segments that encode products which are otherwise lacking in the recipient cell (*e.g.*, tRNA genes, auxotrophic markers); (3) DNA segments that encode products which suppress the activity of a gene product; (4) DNA segments that encode products which can be readily identified (*e.g.*, phenotypic markers such as  $\beta$ -galactosidase, green fluorescent protein (GFP), and cell surface proteins); (5) DNA segments that bind products which are otherwise detrimental to cell survival and/or function; (6) DNA segments that otherwise inhibit the activity of any of the DNA segments described in Nos. 1-5 above (*e.g.*, antisense oligonucleotides); (7) DNA segments that bind products that modify a substrate (*e.g.* restriction endonucleases); (8) DNA segments that can be used to isolate or identify a desired molecule (*e.g.* specific protein binding sites); (9) DNA segments that encode a specific nucleotide sequence which can be otherwise non-functional (*e.g.*, for PCR amplification of subpopulations of molecules); (10) DNA segments, which when absent, directly or indirectly confer resistance or sensitivity to particular compounds; and/or (11) DNA segments that encode products which are toxic in recipient cells.

**Selection scheme:** is any method which allows selection, enrichment, or identification of a desired Product or Product(s) from a mixture containing the Insert Donor, Vector Donor, any intermediates (*e.g.* a Cointegrate), and/or Byproducts. The selection schemes of one preferred embodiment have at least two components that are either linked or unlinked during recombinational cloning. One component is a Selectable marker. The other component controls the expression *in vitro* or *in vivo* of the Selectable marker, or survival of the cell harboring the plasmid carrying the Selectable marker. Generally, this controlling element will be a repressor or inducer of the Selectable marker, but other means for controlling expression of the Selectable marker can be used. Whether a repressor or activator is used will depend on whether the marker is for a positive



or negative selection, and the exact arrangement of the various DNA segments, as will be readily apparent to those skilled in the art. A preferred requirement is that the selection scheme results in selection of or enrichment for only one or more desired Products. As defined herein, selecting for a DNA molecule includes (a) selecting or enriching for the presence of the desired DNA molecule, and (b) selecting or enriching against the presence of DNA molecules that are not the desired DNA molecule.

In one embodiment, the selection schemes (which can be carried out in reverse) will take one of three forms, which will be discussed in terms of Figure 1. The first, exemplified herein with a Selectable marker and a repressor therefore, selects for molecules having segment *D* and lacking segment *C*. The second selects against molecules having segment *C* and for molecules having segment *D*. Possible embodiments of the second form would have a DNA segment carrying a gene toxic to cells into which the *in vitro* reaction products are to be introduced. A toxic gene can be a DNA that is expressed as a toxic gene product (a toxic protein or RNA), or can be toxic in and of itself. (In the latter case, the toxic gene is understood to carry its classical definition of "heritable trait".)

Examples of such toxic gene products are well known in the art, and include, but are not limited to, restriction endonucleases (*e.g.*, *DpnI*), apoptosis-related genes (*e.g.* ASK1 or members of the *bcl-2/ced-9* family), retroviral genes including those of the human immunodeficiency virus (HIV), defensins such as NP-1, inverted repeats or paired palindromic DNA sequences, bacteriophage lytic genes such as those from  $\phi$ X174 or bacteriophage T4; antibiotic sensitivity genes such as *rpsL*, antimicrobial sensitivity genes such as *pheS*, plasmid killer genes, eukaryotic transcriptional vector genes that produce a gene product toxic to bacteria, such as GATA-1, and genes that kill hosts in the absence of a suppressing function, *e.g.*, *kicB* or *ccdB*. A toxic gene can alternatively be selectable *in vitro*, *e.g.*, a restriction site.

Many genes coding for restriction endonucleases operably linked to inducible promoters are known, and may be used in the present invention. *See, e.g.* U.S. Patent Nos. 4,960,707 (*DpnI* and *DpnII*); 5,000,333, 5,082,784 and 5,192,675 (*KpnI*); 5,147,800 (*NgoAIII* and *NgoAI*); 5,179,015 (*FspI* and *HaeIII*):

5,200,333 (*Hae*II and *Taq*I); 5,248,605 (*Hpa*II); 5,312,746 (*Cla*I); 5,231,021 and 5,304,480 (*Xho*I and *Xho*II); 5,334,526 (*Alu*I); 5,470,740 (*Nsi*I); 5,534,428 (*Sst*I/*Sac*I); 5,202,248 (*Nco*I); 5,139,942 (*Nde*I); and 5,098,839 (*Pac*I). *See also* Wilson, G.G., *Nucl. Acids Res.* 19:2539-2566 (1991); and Lunnen, K.D., *et al.*, *Gene* 74:25-32 (1988).

In the second form, segment *D* carries a Selectable marker. The toxic gene would eliminate transformants harboring the Vector Donor, Cointegrate, and Byproduct molecules, while the Selectable marker can be used to select for cells containing the Product and against cells harboring only the Insert Donor.

The third form selects for cells that have both segments *A* and *D* in *cis* on the same molecule, but not for cells that have both segments in *trans* on different molecules. This could be embodied by a Selectable marker that is split into two inactive fragments, one each on segments *A* and *D*.

The fragments are so arranged relative to the recombination sites that when the segments are brought together by the recombination event, they reconstitute a functional Selectable marker. For example, the recombinational event can link a promoter with a structural gene, can link two fragments of a structural gene, or can link genes that encode a heterodimeric gene product needed for survival, or can link portions of a replicon.

**Site-specific recombinase:** is a type of recombinase which typically has at least the following four activities (or combinations thereof): (1) recognition of one or two specific nucleic acid sequences; (2) cleavage of said sequence or sequences; (3) topoisomerase activity involved in strand exchange; and (4) ligase activity to reseal the cleaved strands of nucleic acid. *See* Sauer, B., *Current Opinions in Biotechnology* 5:521-527 (1994). Conservative site-specific recombination is distinguished from homologous recombination and transposition by a high degree of specificity for both partners. The strand exchange mechanism involves the cleavage and rejoining of specific DNA sequences in the absence of DNA synthesis (Landy, A. (1989) *Ann. Rev. Biochem.* 58:913-949).

**Subcloning vector:** is a cloning vector comprising a circular or linear nucleic acid molecule which includes preferably an appropriate replicon. In the

present invention, the subcloning vector (segment **D** in Figure 1) can also contain functional and/or regulatory elements that are desired to be incorporated into the final product to act upon or with the cloned DNA Insert (segment **A** in Figure 1). The subcloning vector can also contain a Selectable marker (preferably DNA).

**Vector:** is a nucleic acid molecule (preferably DNA) that provides a useful biological or biochemical property to an Insert. Examples include plasmids, phages, autonomously replicating sequences (ARS), centromeres, and other sequences which are able to replicate or be replicated *in vitro* or in a host cell, or to convey a desired nucleic acid segment to a desired location within a host cell. A Vector can have one or more restriction endonuclease recognition sites at which the sequences can be cut in a determinable fashion without loss of an essential biological function of the vector, and into which a nucleic acid fragment can be spliced in order to bring about its replication and cloning. Vectors can further provide primer sites, *e.g.*, for PCR, transcriptional and/or translational initiation and/or regulation sites, recombinational signals, replicons, Selectable markers, *etc.* Clearly, methods of inserting a desired nucleic acid fragment which do not require the use of homologous recombination, transpositions or restriction enzymes (such as, but not limited to, UDG cloning of PCR fragments (U.S. Patent No. 5,334,575, entirely incorporated herein by reference), T:A cloning, and the like) can also be applied to clone a fragment into a cloning vector to be used according to the present invention. The cloning vector can further contain one or more selectable markers suitable for use in the identification of cells transformed with the cloning vector.

**Vector Donor:** is one of the two parental nucleic acid molecules (*e.g.* RNA or DNA) of the present invention which carries the DNA segments comprising the DNA vector which is to become part of the desired Product. The Vector Donor comprises a subcloning vector **D** (or it can be called the cloning vector if the Insert Donor does not already contain a cloning vector) and a segment **C** flanked by recombination sites (see Figure 1). Segments **C** and/or **D** can contain elements that contribute to selection for the desired Product daughter molecule, as described above for selection schemes. The recombination signals

can be the same or different, and can be acted upon by the same or different recombinases. In addition, the Vector Donor can be linear or circular.

**Primer:** refers to a single stranded or double stranded oligonucleotide that is extended by covalent bonding of nucleotide monomers during amplification or polymerization of a nucleic acid molecule (e.g. a DNA molecule). In a preferred aspect, the primer comprises one or more recombination sites or portions of such recombination sites. Portions of recombination sites comprise at least 2 bases, at least 5 bases, at least 10 bases or at least 20 bases of the recombination sites of interest. When using portions of recombination sites, the missing portion of the recombination site may be provided by the newly synthesized nucleic acid molecule. Such recombination sites may be located within and/or at one or both termini of the primer. Preferably, additional sequences are added to the primer adjacent to the recombination site(s) to enhance or improve recombination and/or to stabilize the recombination site during recombination. Such stabilization sequences may be any sequences (preferably G/C rich sequences) of any length. Preferably, such sequences range in size from 1 to about 1000 bases, 1 to about 500 bases, and 1 to about 100 bases, 1 to about 60 bases, 1 to about 25, 1 to about 10, 2 to about 10 and preferably about 4 bases. Preferably, such sequences are greater than 1 base in length and preferably greater than 2 bases in length.

**Template:** refers to double stranded or single stranded nucleic acid molecules which are to be amplified, synthesized or sequenced. In the case of double stranded molecules, denaturation of its strands to form a first and a second strand is preferably performed before these molecules will be amplified, synthesized or sequenced, or the double stranded molecule may be used directly as a template. For single stranded templates, a primer complementary to a portion of the template is hybridized under appropriate conditions and one or more polypeptides having polymerase activity (e.g. DNA polymerases and/or reverse transcriptases) may then synthesize a nucleic acid molecule complementary to all or a portion of said template. Alternatively, for double stranded templates, one or more promoters may be used in combination with one or more polymerases to make nucleic acid molecules complementary to all or a portion of the template. The newly synthesized molecules, according to the invention, may be equal or

shorter in length than the original template. Additionally, a population of nucleic acid templates may be used during synthesis or amplification to produce a population of nucleic acid molecules typically representative of the original template population.

5           **Adapter:** is an oligonucleotide or nucleic acid fragment or segment (preferably DNA) which comprises one or more recombination sites (or portions of such recombination sites) which in accordance with the invention can be added to a circular or linear Insert Donor molecule as well as other nucleic acid molecules described herein. When using portions of recombination sites, the missing portion may be provided by the Insert Donor molecule. Such adapters may be added at any location within a circular or linear molecule, although the adapters are preferably added at or near one or both termini of a linear molecule. Preferably, adapters are positioned to be located on both sides (flanking) a particularly nucleic acid molecule of interest. In accordance with the invention, adapters may be added to nucleic acid molecules of interest by standard recombinant techniques (e.g. restriction digest and ligation). For example, adapters may be added to a circular molecule by first digesting the molecule with an appropriate restriction enzyme, adding the adapter at the cleavage site and reforming the circular molecule which contains the adapter(s) at the site of cleavage. Alternatively, adapters may be ligated directly to one or more and preferably both termini of a linear molecule thereby resulting in linear molecule(s) having adapters at one or both termini. In one aspect of the invention, adapters may be added to a population of linear molecules, (e.g. a cDNA library or genomic DNA which has been cleaved or digested) to form a population of linear molecules containing adapters at one and preferably both termini of all or substantial portion of said population.

20           **Library:** refers to a collection of nucleic acid molecules (circular or linear). In one preferred embodiment, a library is representative of all or a significant portion of the DNA content of an organism (a "genomic" library), or a set of nucleic acid molecules representative of all or a significant portion of the expressed genes (a cDNA library) in a cell, tissue, organ or organism. A library may also comprise random sequences made by *de novo* synthesis, mutagenesis of

one or more sequences and the like. Such libraries may or may not be contained in one or more vectors.

**Amplification:** refers to any *in vitro* method for increasing a number of copies of a nucleotide sequence with the use of a polymerase. Nucleic acid amplification results in the incorporation of nucleotides into a DNA and/or RNA molecule or primer thereby forming a new molecule complementary to a template. The formed nucleic acid molecule and its template can be used as templates to synthesize additional nucleic acid molecules. As used herein, one amplification reaction may consist of many rounds of replication. DNA amplification reactions include, for example, polymerase chain reaction (PCR). One PCR reaction may consist of 5-100 "cycles" of denaturation and synthesis of a DNA molecule.

**Oligonucleotide:** refers to a synthetic or natural molecule comprising a covalently linked sequence of nucleotides which are joined by a phosphodiester bond between the 3' position of the deoxyribose or ribose of one nucleotide and the 5' position of the deoxyribose or ribose of the adjacent nucleotide.

**Nucleotide:** refers to a base-sugar-phosphate combination. Nucleotides are monomeric units of a nucleic acid sequence (DNA and RNA). The term nucleotide includes ribonucleoside triphosphates ATP, UTP, CTP, GTP and deoxyribonucleoside triphosphates such as dATP, dCTP, dITP, dUTP, dGTP, dTTP, or derivatives thereof. Such derivatives include, for example, [ $\alpha$ S]dATP, 7-deaza-dGTP and 7-deaza-dATP. The term nucleotide as used herein also refers to dideoxyribonucleoside triphosphates (ddNTPs) and their derivatives. Illustrated examples of dideoxyribonucleoside triphosphates include, but are not limited to, ddATP, ddCTP, ddGTP, ddITP, and ddTTP. According to the present invention, a "nucleotide" may be unlabeled or detectably labeled by well known techniques. Detectable labels include, for example, radioactive isotopes, fluorescent labels, chemiluminescent labels, bioluminescent labels and enzyme labels.

**Hybridization:** The terms "hybridization" and "hybridizing" refers to base pairing of two complementary single-stranded nucleic acid molecules (RNA and/or DNA) to give a double stranded molecule. As used herein, two nucleic acid molecules may be hybridized, although the base pairing is not completely

complementary. Accordingly, mismatched bases do not prevent hybridization of two nucleic acid molecules provided that appropriate conditions, well known in the art, are used.

Other terms used in the fields of recombinant DNA technology and molecular and cell biology as used herein will be generally understood by one of ordinary skill in the applicable arts.

### ***Recombination Schemes***

One general scheme for an *in vitro* or *in vivo* method of the invention is shown in Figure 1, where the Insert Donor and the Vector Donor can be either circular or linear DNA, but is shown as circular. Vector **D** is exchanged for the original cloning vector **B**. The Insert Donor need not comprise a vector. The method of the invention allows the Inserts **A** to be transferred into any number of vectors. According to the invention, the Inserts may be transferred to a particular Vector or may be transferred to a number of vectors in one step. Additionally, the Inserts may be transferred to any number of vectors sequentially, for example, by using the Product DNA molecule as the Insert Donor in combination with a different Vector Donor. The nucleic acid molecule of interest may be transferred into a new vector thereby producing a new Product DNA molecule. The new Product DNA molecule may then be used as starting material to transfer the nucleic acid molecule of interest into a new vector. Such sequential transfers can be performed a number of times in any number of different vectors. Thus the invention allows for cloning or subcloning nucleic acid molecules and because of the ease and simplicity, these methods are particularly suited for high through-put applications. In accordance with the invention, it is desirable to select for the daughter molecule containing elements **A** and **D** and against other molecules, including one or more Cointegrate(s). The square and circle are different sets of recombination sites (e.g., *lox* sites or *att* sites). Segment **A** or **D** can contain at least one Selection Marker, expression signals, origins of replication, or specialized functions for detecting, selecting, expressing, mapping or sequencing DNA, where **D** is used in this example. This scheme can also be reversed

according to the present invention, as described herein. The resulting product of the reverse reaction (e.g. the Insert Donor) may then be used in combination with one or a number of vectors to produce new product molecules in which the Inserts are contained by any number of vectors.

5           Examples of desired DNA segments that can be part of Element *A* or *D* include, but are not limited to, PCR products, large DNA segments, genomic clones or fragments, cDNA clones or fragments, functional elements, *etc.*, and genes or partial genes, which encode useful nucleic acids or proteins. Moreover, the recombinational cloning of the present invention can be used to make *ex vivo* and *in vivo* gene transfer vehicles for protein expression (native or fusion proteins) and/or gene therapy.

10           In Figure 1, the scheme provides the desired Product as containing *A* and Vector *D*, as follows. The Insert Donor (containing *A* and *B*) is first recombined at the square recombination sites by recombination proteins, with the Vector Donor (containing *C* and *D*), to form a Co-integrate having each of *A-D-C-B*. Next, recombination occurs at the circle recombination sites to form Product DNA (*A* and *D*) and Byproduct DNA (*C* and *B*). However, if desired, two or more different Co-integrates can be formed to generate two or more Products.

15           In one embodiment of the present *in vitro* or *in vivo* recombinational cloning method, a method for selecting at least one desired Product DNA is provided. This can be understood by consideration of the map of plasmid pEZC726 depicted in Figure 2. The two exemplary recombination sites are *attP* and *loxP*. On one segment defined by these sites is a kanamycin resistance gene whose promoter has been replaced by the *tetOP* operator/promoter from transposon *Tn10*. In the absence of tet repressor protein, *E. coli* RNA polymerase transcribes the kanamycin resistance gene from the *tetOP*. If tet repressor is present, it binds to *tetOP* and blocks transcription of the kanamycin resistance gene. The other segment of pEZC726 has the *tet* repressor gene expressed by a constitutive promoter. Thus cells transformed by pEZC726 are resistant to chloramphenicol, because of the chloramphenicol acetyl transferase gene on the same segment as *tetR*, but are sensitive to kanamycin. The recombination

20

25

30



reactions result in separation of the *tetR* gene from the regulated kanamycin resistance gene. This separation results in kanamycin resistance in cells receiving the desired recombination Product.

Two different sets of plasmids were constructed to demonstrate the *in vitro* method. One set, for use with Cre recombinase only (cloning vector 602 and subcloning vector 629 (Figure 3)) contained *loxP* and *loxP 511* sites. A second set, for use with Cre and integrase (cloning vector 705 and subcloning vector 726 (Figure 2)) contained *loxP* and *att* sites. The efficiency of production of the desired daughter plasmid was about 60 fold higher using both enzymes than using Cre alone. Nineteen of twenty four colonies from the Cre-only reaction contained the desired product, while thirty eight of thirty eight colonies from the integrase plus Cre reaction contained the desired product plasmid.

A variety of other selection schemes can be used that are known in the art as they can suit a particular purpose for which the recombinational cloning is carried out. Depending upon individual preferences and needs, a number of different types of selection schemes can be used in the recombinational cloning or subcloning methods of the present invention. The skilled artisan can take advantage of the availability of the many DNA segments or methods for making them and the different methods of selection that are routinely used in the art. Such DNA segments include but are not limited to those which encodes an activity such as, but not limited to, production of RNA, peptide, or protein, or providing a binding site for such RNA, peptide, or protein. Examples of DNA molecules used in devising a selection scheme are given above, under the definition of "selection scheme"

Additional examples include but are not limited to:

- (i) Generation of new primer sites for PCR (*e.g.*, juxtaposition of two DNA sequences that were not previously juxtaposed);
- (ii) Inclusion of a DNA sequence acted upon by a restriction endonuclease or other DNA modifying enzyme, chemical, ribozyme, etc.;
- (iii) Inclusion of a DNA sequence recognized by a DNA binding protein, RNA, DNA, chemical, etc.) (*e.g.*, for use as an affinity tag

for selecting for or excluding from a population (Davis, *Nucl. Acids Res.* 24:702-706 (1996); *J. Virol.* 69: 8027-8034 (1995)) or for juxtaposing a promoter for *in vitro* transcription;

- (iv) *In vitro* selection of RNA ligands for the ribosomal L22 protein associated with Epstein-Barr virus-expressed RNA by using randomized and cDNA-derived RNA libraries;
- (vi) The positioning of functional elements whose activity requires a specific orientation or juxtaposition (*e.g.*, (a) a recombination site which reacts poorly in *trans*, but when placed in *cis*, in the presence of the appropriate proteins, results in recombination that destroys certain populations of molecules; (*e.g.*, reconstitution of a promoter sequence that allows *in vitro* RNA synthesis). The RNA can be used directly, or can be reverse transcribed to obtain the desired DNA construct;
- (vii) Selection of the desired product by size (*e.g.*, fractionation) or other physical property of the molecule(s); and
- (viii) Inclusion of a DNA sequence required for a specific modification (*e.g.*, methylation) that allows its identification.

After formation of the Product and Byproduct in the method of the present invention, the selection step can be carried out either *in vitro* or *in vivo* depending upon the particular selection scheme which has been optionally devised in the particular recombinational cloning procedure.

For example, an *in vitro* method of selection can be devised for the Insert Donor and Vector Donor DNA molecules. Such scheme can involve engineering a rare restriction site in the starting circular vectors in such a way that after the recombination events the rare cutting sites end up in the Byproduct. Hence, when the restriction enzyme which binds and cuts at the rare restriction site is added to the reaction mixture *in vitro*, all of the DNA molecules carrying the rare cutting site, *i.e.*, the starting DNA molecules, the Cointegrate, and the Byproduct, will be cut and rendered nonreplicable in the intended host cell. For example, cutting sites in segments **B** and **C** (see Figure 1) can be used to select against all molecules

except the Product. Alternatively, only a cutting site in *C* is needed if one is able to select for segment *D*, *e.g.*, by a drug resistance gene not found on *B*.

Similarly, an *in vitro* selection method can be devised when dealing with linear DNA molecules. DNA sequences complementary to a PCR primer sequence can be so engineered that they are transferred, through the recombinational cloning method, only to the Product molecule. After the reactions are completed, the appropriate primers are added to the reaction solution and the sample is subjected to PCR. Hence, all or part of the Product molecule is amplified.

Other *in vivo* selection schemes can be used with a variety of host cells, particularly *E. coli* lines. One is to put a repressor gene on one segment of the subcloning plasmid, and a drug marker controlled by that repressor on the other segment of the same plasmid. Another is to put a killer gene on segment *C* of the subcloning plasmid (Figure 1). Of course a way must exist for growing such a plasmid, *i.e.*, there must exist circumstances under which the killer gene will not kill. There are a number of these genes known which require particular strains of *E. coli*. One such scheme is to use the restriction enzyme *DpnI*, which will not cleave unless its recognition sequence GATC is methylated. Many popular common *E. coli* strains methylate GATC sequences, but there are mutants in which cloned *DpnI* can be expressed without harm. Other restriction enzyme genes may also be used as a toxic gene for selection. In such cases, a host containing a gene encoding the corresponding methylase gene provides protected host for use in the invention. Similarly, the *ccdB* protein is a potent poison of DNA gyrase, efficiently trapping gyrase molecules in a cleavable complex, resulting in DNA strand breakage and cell death. Mutations in the *gyrA* subunit of DNA gyrase, specifically the *gyrA462* mutation, confers resistance to *ccdB* (Bernard and Couturier, J. Mol. Bio. 226 (1992) 735-745). An *E. coli* strain, DB2, has been constructed that contains the *gyrA462* mutation. DB2 cells containing plasmids that express the *ccdB* gene are not killed by *ccdB*. This strain is available from Life Technologies and has been deposited on October 14, 1997 with the Collection, Agricultural Research Culture Collection (NRRL), 1815

North University Street, Peoria, IL 61604 USA as deposit number NRRL B-21852.

Of course analogous selection schemes can be devised for other host organisms. For example, the *tet* repressor/operator of *Tn10* has been adapted to control gene expression in eukaryotes (Gossen, M., and Bujard, H., *Proc. Natl. Acad. Sci. USA* 89:5547-5551 (1992)). Thus the same control of drug resistance by the *tet* repressor exemplified herein or other selection schemes described herein can be applied to select for Product in eukaryotic cells.

### ***Recombination Proteins***

In the present invention, the exchange of DNA segments is achieved by the use of recombination proteins, including recombinases and associated co-factors and proteins. Various recombination proteins are described in the art. Examples of such recombinases include:

**Cre:** A protein from bacteriophage P1 (Abremski and Hoess, *J. Biol. Chem.* 259(3):1509-1514 (1984)) catalyzes the exchange (i.e., causes recombination) between 34 bp DNA sequences called *loxP* (locus of crossover) sites (See Hoess *et al.*, *Nucl. Acids Res.* 14(5):2287 (1986)). Cre is available commercially (Novagen, Catalog No. 69247-1). Recombination mediated by Cre is freely reversible. From thermodynamic considerations it is not surprising that Cre-mediated integration (recombination between two molecules to form one molecule) is much less efficient than Cre-mediated excision (recombination between two *loxP* sites in the same molecule to form two daughter molecules). Cre works in simple buffers with either magnesium or spermidine as a cofactor, as is well known in the art. The DNA substrates can be either linear or supercoiled. A number of mutant *loxP* sites have been described (Hoess *et al.*, *supra*). One of these, *loxP 511*, recombines with another *loxP 511* site, but will not recombine with a *loxP* site.

**Integrase:** A protein from bacteriophage lambda that mediates the integration of the lambda genome into the *E. coli* chromosome. The bacteriophage  $\lambda$  Int recombinational proteins promote recombination between its substrate *att* sites as part of the formation or induction of a lysogenic state.

Reversibility of the recombination reactions results from two independent pathways for integrative and excisive recombination. Each pathway uses a unique, but overlapping, set of the 15 protein binding sites that comprise *att* site DNAs. Cooperative and competitive interactions involving four proteins (Int, Xis, IHF and FIS) determine the direction of recombination.

Integrative recombination involves the Int and IHF proteins and sites *attP* (240 bp) and *attB* (25 bp). Recombination results in the formation of two new sites: *attL* and *attR*. Excisive recombination requires Int, IHF, and Xis, and sites *attL* and *attR* to generate *attP* and *attB*. Under certain conditions, FIS stimulates excisive recombination. In addition to these normal reactions, it should be appreciated that *attP* and *attB*, when placed on the same molecule, can promote *excisive* recombination to generate two excision products, one with *attL* and one with *attR*. Similarly, intermolecular recombination between molecules containing *attL* and *attR*, in the presence of Int, IHF and Xis, can result in *integrative* recombination and the generation of *attP* and *attB*. Hence, by flanking DNA segments with appropriate combinations of engineered *att* sites, in the presence of the appropriate recombination proteins, one can direct excisive or integrative recombination, as reverse reactions of each other.

Each of the *att* sites contains a 15 bp core sequence; individual sequence elements of functional significance lie within, outside, and across the boundaries of this common core (Landy, A., *Ann. Rev. Biochem.* 58:913 (1989)). Efficient recombination between the various *att* sites requires that the sequence of the central common region be identical between the recombining partners, however, the exact sequence is now found to be modifiable. Consequently, derivatives of the *att* site with changes within the core are now discovered to recombine as least as efficiently as the native core sequences.

Integrase acts to recombine the *attP* site on bacteriophage lambda (about 240 bp) with the *attB* site on the *E. coli* genome (about 25 bp) (Weisberg, R.A. and Landy, A. in *Lambda II*, p. 211 (1983), Cold Spring Harbor Laboratory)), to produce the integrated lambda genome flanked by *attL* (about 100 bp) and *attR* (about 160 bp) sites. In the absence of Xis (see below), this reaction is essentially irreversible. The integration reaction mediated by integrase and IHF works

*in vitro*, with simple buffer containing spermidine. Integrase can be obtained as described by Nash, H.A., *Methods of Enzymology* 100:210-216 (1983). IHF can be obtained as described by Filutowicz, M., *et al.*, *Gene* 147:149-150 (1994).

Numerous recombination systems from various organisms can also be used, based on the teaching and guidance provided herein. *See, e.g.*, Hoess *et al.*, *Nucleic Acids Research* 14(6):2287 (1986); Abremski *et al.*, *J. Biol. Chem.* 261(1):391 (1986); Campbell, *J. Bacteriol.* 174(23):7495 (1992); Qian *et al.*, *J. Biol. Chem.* 267(11):7794 (1992); Araki *et al.*, *J. Mol. Biol.* 225(1):25 (1992)). Many of these belong to the integrase family of recombinases (Argos *et al.* *EMBO J.* 5:433-440 (1986)). Perhaps the best studied of these are the Integrase/*att* system from bacteriophage  $\lambda$  (Landy, A. (1993) *Current Opinions in Genetics and Devel.* 3:699-707), the Cre/*loxP* system from bacteriophage P1 (Hoess and Abremski (1990) In *Nucleic Acids and Molecular Biology*, vol. 4. Eds.: Eckstein and Lilley, Berlin-Heidelberg: Springer-Verlag; pp. 90-109), and the FLP/FRT system from the *Saccharomyces cerevisiae* 2  $\mu$  circle plasmid (Broach *et al.* *Cell* 29:227-234 (1982)).

Members of a second family of site-specific recombinases, the resolvase family (*e.g.*,  $\gamma\delta$ , Tn3 resolvase, Hin, Gin, and Cin) are also known. Members of this highly related family of recombinases are typically constrained to intramolecular reactions (*e.g.*, inversions and excisions) and can require host-encoded factors. Mutants have been isolated that relieve some of the requirements for host factors (Maeser and Kahnmann (1991) *Mol. Gen. Genet.* 230:170-176), as well as some of the constraints of intramolecular recombination.

Other site-specific recombinases similar to  $\lambda$  Int and similar to P1 Cre can be substituted for Int and Cre. Such recombinases are known. In many cases the purification of such other recombinases has been described in the art. In cases when they are not known, cell extracts can be used or the enzymes can be partially purified using procedures described for Cre and Int.

While Cre and Int are described in detail for reasons of example, many related recombinase systems exist and their application to the described invention is also provided according to the present invention. The integrase family of site-specific recombinases can be used to provide alternative recombination proteins

and recombination sites for the present invention, as site-specific recombination proteins encoded by, for example bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1. This group of proteins exhibits an unexpectedly large diversity of sequences. Despite this diversity, all of the recombinases can be aligned in their C-terminal halves. A 40-residue region near the C terminus is particularly well conserved in all the proteins and is homologous to a region near the C terminus of the yeast 2 mu plasmid FIp protein. Three positions are perfectly conserved within this family: histidine, arginine and tyrosine are found at respective alignment positions 396, 399 and 433 within the well-conserved C-terminal region. These residues contribute to the active site of this family of recombinases, and suggest that tyrosine-433 forms a transient covalent linkage to DNA during strand cleavage and rejoining. See, e.g., Argos, P. *et al.*, *EMBO J.* 5:433-40 (1986).

The recombinases of some transposons, such as those of conjugative transposons (e.g., Tn916) (Scott and Churchward. 1995. *Ann Rev Microbiol* 49:367; Taylor and Churchward, 1997. *J Bacteriol* 179:1837) belong to the integrase family of recombinases and in some cases show strong preferences for specific integration sites (Ike et al 1992. *J Bacteriol* 174:1801; Trieu-Cuot et al, 1993. *Mol. Microbiol* 8:179).

Alternatively, IS231 and other *Bacillus thuringiensis* transposable elements could be used as recombination proteins and recombination sites. *Bacillus thuringiensis* is an entomopathogenic bacterium whose toxicity is due to the presence in the sporangia of delta-endotoxin crystals active against agricultural pests and vectors of human and animal diseases. Most of the genes coding for these toxin proteins are plasmid-borne and are generally structurally associated with insertion sequences (IS231, IS232, IS240, ISBT1 and ISBT2) and transposons (Tn4430 and Tn5401). Several of these mobile elements have been shown to be active and participate in the crystal gene mobility, thereby contributing to the variation of bacterial toxicity.

Structural analysis of the iso-IS231 elements indicates that they are related to IS1151 from *Clostridium perfringens* and distantly related to IS4 and IS186 from *Escherichia coli*. Like the other IS4 family members, they contain a conserved transposase-integrase motif found in other IS families and retroviruses.

Moreover, functional data gathered from IS231A in *Escherichia coli* indicate a non-replicative mode of transposition, with a preference for specific targets. Similar results were also obtained in *Bacillus subtilis* and *B. thuringiensis*. See, e.g., Mahillon, J. *et al.*, *Genetica* 93:13-26 (1994); Campbell, J. *Bacteriol.* 7495-7499 (1992).

An unrelated family of recombinases, the transposases, have also been used to transfer genetic information between replicons. Transposons are structurally variable, being described as simple or compound, but typically encode the recombinase gene flanked by DNA sequences organized in inverted orientations. Integration of transposons can be random or highly specific. Representatives such as Tn7, which are highly site-specific, have been applied to the efficient movement of DNA segments between replicons (Lucklow et al. 1993. *J. Virol* 67:4566-4579).

A related element, the integron, are also translocatable-promoting movement of drug resistance cassettes from one replicon to another. Often these elements are defective transposon derivatives. Transposon Tn21 contains a class I integron called In2. The integrase (IntI1) from In2 is common to all integrons in this class and mediates recombination between two 59-bp elements or between a 59-bp element and an attI site that can lead to insertion into a recipient integron. The integrase also catalyzes excisive recombination. (Hall, 1997. *Ciba Found Symp* 207:192; Francia et al., 1997. *J Bacteriol* 179:4419).

Group II introns are mobile genetic elements encoding a catalytic RNA and protein. The protein component possesses reverse transcriptase, maturase and an endonuclease activity, while the RNA possesses endonuclease activity and determines the sequence of the target site into which the intron integrates. By modifying portions of the RNA sequence, the integration sites into which the element integrates can be defined. Foreign DNA sequences can be incorporated between the ends of the intron, allowing targeting to specific sites. This process, termed retrohoming, occurs via a DNA:RNA intermediate, which is copied into cDNA and ultimately into double stranded DNA (Matsuura et al., *Genes and Dev* 1997; Guo et al, *EMBO J*, 1997). Numerous intron-encoded homing endonucleases have been identified (Belfort and Roberts, 1997. *NAR*



25:3379). Such systems can be easily adopted for application to the described subcloning methods.

The amount of recombinase which is added to drive the recombination reaction can be determined by using known assays. Specifically, titration assay is used to determine the appropriate amount of a purified recombinase enzyme, or the appropriate amount of an extract.

### ***Engineered Recombination Sites***

The above recombinases and corresponding recombinase sites are suitable for use in recombination cloning according to the present invention. However, wild-type recombination sites may contain sequences that reduce the efficiency or specificity of recombination reactions or the function of the Product molecules as applied in methods of the present invention. For example, multiple stop codons in attB, attR, attP, attL and loxP recombination sites occur in multiple reading frames on both strands, so translation efficiencies are reduced, *e.g.*, where the coding sequence must cross the recombination sites, (only one reading frame is available on each strand of loxP and attB sites) or impossible (in attP, attR or attL).

Accordingly, the present invention also provides engineered recombination sites that overcome these problems. For example, att sites can be engineered to have one or multiple mutations to enhance specificity or efficiency of the recombination reaction and the properties of Product DNAs (*e.g.*, att1, att2, and att3 sites); to decrease reverse reaction (*e.g.*, removing P1 and H1 from attR). The testing of these mutants determines which mutants yield sufficient recombinational activity to be suitable for recombination subcloning according to the present invention.

Mutations can therefore be introduced into recombination sites for enhancing site specific recombination. Such mutations include, but are not limited to: recombination sites without translation stop codons that allow fusion proteins to be encoded; recombination sites recognized by the same proteins but differing in base sequence such that they react largely or exclusively with their homologous partners allowing multiple reactions to be contemplated; and mutations that

prevent hairpin formation of recombination sites. Which particular reactions take place can be specified by which particular partners are present in the reaction mixture. For example, a tripartite protein fusion could be accomplished with parental plasmids containing recombination sites attR1 and attL1; and attB3; attR1; attP3 and 10xP; and/or attR3 and 10xP; and/or attR3 and attL2.

There are well known procedures for introducing specific mutations into nucleic acid sequences. A number of these are described in Ausubel, F.M. *et al.*, *Current Protocols in Molecular Biology*, Wiley Interscience, New York (1989-1996). Mutations can be designed into oligonucleotides, which can be used to modify existing cloned sequences, or in amplification reactions. Random mutagenesis can also be employed if appropriate selection methods are available to isolate the desired mutant DNA or RNA. The presence of the desired mutations can be confirmed by sequencing the nucleic acid by well known methods.

The following non-limiting methods can be used to modify or mutate a core region of a given recombination site to provide mutated sites that can be used in the present invention:

1. By recombination of two parental DNA sequences by site-specific (e.g. attL and attR to give attB) or other (e.g. homologous) recombination mechanisms where the parental DNA segments contain one or more base alterations resulting in the final mutated core sequence;
2. By mutation or mutagenesis (site-specific, PCR, random, spontaneous, etc) directly of the desired core sequence;
3. By mutagenesis (site-specific, PCR, random, spontaneous, etc) of parental DNA sequences, which are recombined to generate a desired core sequence;
4. By reverse transcription of an RNA encoding the desired core sequence; and
5. By de novo synthesis (chemical synthesis) of a sequence having the desired base changes.

The functionality of the mutant recombination sites can be demonstrated in ways that depend on the particular characteristic that is desired. For example, the lack of translation stop codons in a recombination site can be demonstrated by expressing the appropriate fusion proteins. Specificity of recombination between homologous partners can be demonstrated by introducing the appropriate molecules into *in vitro* reactions, and assaying for recombination products as described herein or known in the art. Other desired mutations in recombination sites might include the presence or absence of restriction sites, translation or transcription start signals, protein binding sites, and other known functionalities of nucleic acid base sequences. Genetic selection schemes for particular functional attributes in the recombination sites can be used according to known method steps. For example, the modification of sites to provide (from a pair of sites that do not interact) partners that do interact could be achieved by requiring deletion, via recombination between the sites, of a DNA sequence encoding a toxic substance. Similarly, selection for sites that remove translation stop sequences, the presence or absence of protein binding sites, etc., can be easily devised by those skilled in the art.

Accordingly, the present invention provides a nucleic acid molecule, comprising at least one DNA segment having at least two engineered recombination sites flanking a Selectable marker and/or a desired DNA segment, wherein at least one of said recombination sites comprises a core region having at least one engineered mutation that enhances recombination *in vitro* in the formation of a Cointegrate DNA or a Product DNA.

While in the preferred embodiment the recombination sites differ in sequence and do not interact with each other, it is recognized that sites comprising the same sequence can be manipulated to inhibit recombination with each other. Such conceptions are considered and incorporated herein. For example, a protein binding site can be engineered adjacent to one of the sites. In the presence of the protein that recognizes said site, the recombinase fails to access the site and the other site is therefore used preferentially. In the cointegrate this site can no longer react since it has been changed e.g. from attB to attL. In resolution of the

cointegrate, the protein can be inactivated (e.g. by antibody, heat or a change of buffer) and the second site can undergo recombination.

The nucleic acid molecule can have at least one mutation that confers at least one enhancement of said recombination, said enhancement selected from the group consisting of substantially (i) favoring integration; (ii) favoring recombination; (iii) relieving the requirement for host factors; (iii) increasing the efficiency of said Cointegrate DNA or Product DNA formation; and (iv) increasing the specificity of said Cointegrate DNA or Product DNA formation.

The nucleic acid molecule preferably comprises at least one recombination site derived from attB, attP, attL or attR, such as attR' or attP'. More preferably the att site is selected from att1, att2, or att3, as described herein.

In a preferred embodiment, the core region comprises a DNA sequence selected from the group consisting of:

- (a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:1);
- (b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:2);
- (c) GTTCAGCTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:3);
- (d) AGCCWGCTTCKTRTACNAAAGTSGB (m-attL) (SEQ ID NO:4);
- (e) GTTCAGCTTTYKTRTACNAAAGTSGB(m-attP1) (SEQ ID NO:5);
- (f) RBYCW GCTTTYTTRTACWAA STKGD (n-att) (SEQ ID NO:39);
- (g) ASCCW GCTTTYTTRTACWAA STKGW (n-attB) (SEQ ID NO:40);
- (h) ASCCW GCTTTYTTRTACWAA GTTGG (n-attL) (SEQ ID NO:41);
- (i) GTTCA GCTTTYTTRTACWAA STKGW (n-attR) (SEQ ID NO:42);
- (j) GTTCA GCTTTYTTRTACWAA GTTGG (n-attP) (SEQ ID NO:43);

or a corresponding or complementary DNA or RNA sequence, wherein R=A or G; K=G or T/U; Y=C or T/U; W=A or T/U; N=A or C or G or T/U; S=C or G; and B=C or G or T/U, as presented in 37 C.F.R. §1.822, which is entirely incorporated herein by reference, wherein the core region does not contain a stop codon in one or more reading frames.

The core region also preferably comprises a DNA sequence selected from the group consisting of:

- (a) AGCCTGCTTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:6);
  - (b) AGCCTGCTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:7);
  - (c) ACCCAGCTTCTTGTACAAAGTGGT (attB3) (SEQ ID NO:8);
  - (d) GTTCAGCTTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:9);
  - (e) GTTCAGCTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:10);
  - (f) GTTCAGCTTCTTGTACAAAGTGGT (attR3) (SEQ ID NO:11);
  - (g) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:12);
  - (h) AGCCTGCTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:13);
  - (i) ACCCAGCTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:14);
  - (j) GTTCAGCTTTTTTGTACAAAGTTGG (attP1) (SEQ ID NO:15);
  - (k) GTTCAGCTTCTTGTACAAAGTTGG (attP2,P3) (SEQ ID NO:16);
- or a corresponding or complementary DNA or RNA sequence.

The present invention thus also provides a method for making a nucleic acid molecule, comprising providing a nucleic acid molecule having at least one engineered recombination site comprising at least one DNA sequence having at least 80-99% homology (or any range or value therein) to at least one of the above sequences, or any suitable recombination site, or which hybridizes under stringent conditions thereto, as known in the art.

Clearly, there are various types and permutations of such well-known *in vitro* and *in vivo* selection methods, each of which are not described herein for the sake of brevity. However, such variations and permutations are contemplated and considered to be the different embodiments of the present invention.

It is important to note that as a result of the preferred embodiment being *in vitro* recombination reactions, non-biological molecules such as PCR products can be manipulated via the present recombinational cloning method. In one example, it is possible to clone linear molecules into circular vectors.

There are a number of applications for the present invention. These uses include, but are not limited to, changing vectors, apposing promoters with genes, constructing genes for fusion proteins, changing copy number, changing replicons, cloning into phages, and cloning, *e.g.*, PCR products (with an *attB* site at one end and a *loxP* site at the other end), genomic DNAs, and cDNAs.

### ***Vector Donors***

In accordance with the invention, any vector may be used to construct the Vector Donors of the invention. In particular, vectors known in the art and those commercially available (and variants or derivatives thereof) may in accordance with the invention be engineered to include one or more recombination sites for use in the methods of the invention. Such vectors may be obtained from, for example, Vector Laboratories Inc., InVitrogen, Promega, Novagen, NEB, Clontech, Boehringer Mannheim, Pharmacia, EpiCenter, OriGenes Technologies Inc., Stratagene, Perkin Elmer, Pharmingen, Life Technologies, Inc., and Research Genetics. Such vectors may then for example be used for cloning or subcloning nucleic acid molecules of interest. General classes of vectors of particular interest include prokaryotic and/or eukaryotic cloning vectors, expression vectors, fusion vectors, two-hybrid or reverse two-hybrid vectors, shuttle vectors for use in different hosts, mutagenesis vectors, transcription vectors, vectors for receiving large inserts and the like.

Other vectors of interest include viral origin vectors (M13 vectors, bacterial phage  $\lambda$  vectors, adenovirus vectors, and retrovirus vectors), high, low and adjustable copy number vectors, vectors which have compatible replicons for

use in combination in a single host (pACYC184 and pBR322) and eukaryotic episomal replication vectors (pCDM8).

Particular vectors of interest include prokaryotic expression vectors such as pcDNA II, pSL301, pSE280, pSE380, pSE420, pTrcHisA, B, and C, pRSET A, B, and C (Invitrogen, Inc.), pGEMEX-1, and pGEMEX-2 (Promega, Inc.), the pET vectors (Novagen, Inc.), pTrc99A, pKK223-3, the pGEX vectors, pEZZ18, pRIT2T, and pMC1871 (Pharmacia, Inc.), pKK233-2 and pKK388-1 (Clontech, Inc.), and pProEx-HT (Life Technologies, Inc.) and variants and derivatives thereof. Vector donors can also be made from eukaryotic expression vectors such as pFastBac, pFastBac HT, pFastBac DUAL, pSFV, and pTet-Splice (Life Technologies, Inc.), pEUK-C1, pPUR, pMAM, pMAMneo, pBI101, pBI121, pDR2, pCMVEBNA, and pYACneo (Clontech), pSVK3, pSVL, pMSG, pCH110, and pKK232-8 (Pharmacia, Inc.), p3'SS, pXT1, pSG5, pPbac, pMbac, pMC1neo, and pOG44 (Stratagene, Inc.), and pYES2, pAC360, pBlueBacHis A, B, and C, pVL1392, pBsueBacIII, pCDM8, pcDNA1, pZeoSV, pcDNA3 pREP4, pCEP4, and pEBVHis (Invitrogen, Inc.) and variants or derivatives thereof.

Other vectors of particular interest include pUC18, pUC19, pBlueScript, pSPORT, cosmid, phagemids, YAC's (yeast artificial chromosomes), BAC's (bacterial artificial chromosomes), P1 (*E. coli* phage), pQE70, pQE60, pQE9 (quagan), pBS vectors, PhageScript vectors, BlueScript vectors, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene), pcDNA3 (Invitrogen), pGEX, pTrsfus, pTrc99A, pET-5, pET-9, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia), pSPORT1, pSPORT2, pCMVSPORT2.0 and pSV-SPORT1 (Life Technologies, Inc.) and variants or derivatives thereof.

Additional vectors of interest include pTrxFus, pThioHis, pLEX, pTrcHis, pTrcHis2, pRSET, pBlueBacHis2, pcDNA3.1/His, pcDNA3.1(-)/Myc-His, pSecTag, pEBVHis, pPIC9K, pPIC3.5K, pAO815, pPICZ, pPICZ $\alpha$ , pGAPZ, pGAPZ $\alpha$ , pBlueBac4.5, pBlueBacHis2, pMelBac, pSinRep5, pSinHis, pIND, pIND(SP1), pVgRXXR, pcDNA2.1, pYES2, pZerO1.1, pZerO-2.1, pCR-Blunt, pSE280, pSE380, pSE420, pVL1392, pVL1393, pCDM8, pcDNA1.1, pcDNA1.1/Amp, pcDNA3.1, pcDNA3.1/Zeo, pSe,SV2, pRc/CMV2, pRc/RSV, pREP4, pREP7, pREP8, pREP9, pREP10, pCEP4, pEBVHis, pCR3.1, pCR2.1,

pCR3.1-Uni, and pCRBac from Invitrogen;  $\lambda$ ExCell,  $\lambda$ gt11, pTrc99A, pKK223-3, pGEX-1 $\lambda$ T, pGEX-2T, pGEX-2TK, pGEX-4T-1, pGEX-4T-2, pGEX-4T-3, pGEX-3X, pGEX-5X-1, pGEX-5X-2, pGEX-5X-3, pEZZ18, pRIT2T, pMC1871, pSVK3, pSVL, pMSG, pCH110, pKK232-8, pSL1180, pNEO, and pUC4K from Pharmacia; pSCREEN-1b(+), pT7Blue(R), pT7Blue-2, pCITE-4abc(+), pOCUS-2, pTag, pET-32 LIC, pET-30 LIC, pBAC-2cp LIC, pBACgus-2cp LIC, pT7Blue-2 LIC, pT7Blue-2,  $\lambda$ SCREEN-1,  $\lambda$ BlueSTAR, pET-3abcd, pET-7abc, pET9abcd, pET11abcd, pET12abc, pET-14b, pET-15b, pET-16b, pET-17b-pET-17xb, pET-19b, pET-20b(+), pET-21abcd(+), pET-22b(+), pET-23abcd(+), pET-24abcd(+), pET-25b(+), pET-26b(+), pET-27b(+), pET-28abc(+), pET-29abc(+), pET-30abc(+), pET-31b(+), pET-32abc(+), pET-33b(+), pBAC-1, pBACgus-1, pBAC4x-1, pBACgus4x-1, pBAC-3cp, pBACgus-2cp, pBACsurf-1, plg, Signal plg, pYX, Selecta Vecta-Neo, Selecta Vecta - Hyg, and Selecta Vecta - Gpt from Novagen; pLexA, pB42AD, pGBT9, pAS2-1, pGAD424, pACT2, pGAD GL, pGAD GH, pGAD10, pGilda, pEZM3, pEGFP, pEGFP-1, pEGFP-N, pEGFP-C, pEBFP, pGFPuv, pGFP, p6xHis-GFP, pSEAP2-Basic, pSEAP2-Contral, pSEAP2-Promoter, pSEAP2-Enhancer, p $\beta$ gal-Basic, p $\beta$ gal-Control, p $\beta$ gal-Promoter, p $\beta$ gal-Enhancer, pCMV $\beta$ , pTet-Off, pTet-On, pTK-Hyg, pRetro-Off, pRetro-On, pIRES1neo, pIRES1hyg, pLXSN, pLNCX, pLAPSN, pMAMneo, pMAMneo-CAT, pMAMneo-LUC, pPUR, pSV2neo, pYEX4T-1/2/3, pYEX-S1, pBacPAK-His, pBacPAK8/9, pAcUW31, BacPAK6, pTriplEx,  $\lambda$ gt10,  $\lambda$ gt11, pWE15, and  $\lambda$ TriplEx from Clontech; Lambda ZAP II, pBK-CMV, pBK-RSV, pBluescript II KS +/-, pBluescript II SK +/-, pAD-GAL4, pBD-GAL4 Cam, pSurfsript, Lambda FIX II, Lambda DASH, Lambda EMBL3, Lambda EMBL4, SuperCos, pCR-Script Amp, pCR-Script Cam, pCR-Script Direct, pBS +/-, pBC KS +/-, pBC SK +/-, Phagescript, pCAL-n-EK, pCAL-n, pCAL-c, pCAL-kc, pET-3abcd, pET-11abcd, pSPUTK, pESP-1, pCMVLacI, pOPRSVI/MCS, pOPI3 CAT, pXT1, pSG5, pPbac, pMbac, pMC1neo, pMC1neo Poly A, pOG44, pOG45, pFRT $\beta$ GAL, pNEO $\beta$ GAL, pRS403, pRS404, pRS405, pRS406, pRS413, pRS414, pRS415, and pRS416 from Stratagene.

Two-hybrid and reverse two-hybrid vectors of particular interest include pPC86, pDBLeu, pDBTrp, pPC97, p2.5, pGAD1-3, pGAD10, pACT, pACT2,



pGADGL, pGADGH, pAS2-1, pGAD424, pGBT8, pGBT9, pGAD-GAL4, pLexA, pBD-GAL4, pHISi, pHISi-1, placZi, pB42AD, pDG202, pJK202, pJG4-5, pNLexA, pYESTrp and variants or derivatives thereof.

## ***Polymerases***

Preferred polypeptides having reverse transcriptase activity (*i.e.*, those polypeptides able to catalyze the synthesis of a DNA molecule from an RNA template) include, but are not limited to Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase, Rous Sarcoma Virus (RSV) reverse transcriptase, Avian Myeloblastosis Virus (AMV) reverse transcriptase, Rous Associated Virus (RAV) reverse transcriptase, Myeloblastosis Associated Virus (MAV) reverse transcriptase, Human Immunodeficiency Virus (HIV) reverse transcriptase, retroviral reverse transcriptase, retrotransposon reverse transcriptase, hepatitis B reverse transcriptase, cauliflower mosaic virus reverse transcriptase and bacterial reverse transcriptase. Particularly preferred are those polypeptides having reverse transcriptase activity that are also substantially reduced in RNase H activity (*i.e.*, "RNase H" polypeptides). By a polypeptide that is "substantially reduced in RNase H activity" is meant that the polypeptide has less than about 20%, more preferably less than about 15%, 10% or 5%, and most preferably less than about 2%, of the RNase H activity of a wildtype or RNase H<sup>+</sup> enzyme such as wildtype M-MLV reverse transcriptase. The RNase H activity may be determined by a variety of assays, such as those described, for example, in U.S. Patent No. 5,244,797, in Kotewicz, M.L. *et al.*, *Nucl. Acids Res.* 16:265 (1988) and in Gerard, G.F., *et al.*, *FOCUS* 14(5):91 (1992), the disclosures of all of which are fully incorporated herein by reference. Suitable RNase H<sup>-</sup> polypeptides for use in the present invention include, but are not limited to, M-MLV H<sup>-</sup> reverse transcriptase, RSV H<sup>-</sup> reverse transcriptase, AMV H<sup>-</sup> reverse transcriptase, RAV H<sup>-</sup> reverse transcriptase, MAV H<sup>-</sup> reverse transcriptase, HIV H<sup>-</sup> reverse transcriptase, and SUPERScript™ I reverse transcriptase and SUPERScript™ II reverse transcriptase which are available commercially, for example from Life Technologies, Inc. (Rockville, Maryland).

Other polypeptides having nucleic acid polymerase activity suitable for use in the present methods include thermophilic DNA polymerases such as DNA polymerase I, DNA polymerase III, Klenow fragment, T7 polymerase, and T5 polymerase, and thermostable DNA polymerases including, but not limited to, *Thermus thermophilus* (*Tth*) DNA polymerase, *Thermus aquaticus* (*Taq*) DNA polymerase, *Thermotoga neopolitana* (*Tne*) DNA polymerase, *Thermotoga maritima* (*Tma*) DNA polymerase, *Thermococcus litoralis* (*Tli* or VENT®) DNA polymerase, *Pyrococcus furiosus* (*Pfu* or DEEPVENT®) DNA polymerase, *Pyrococcus woosii* (*Pwo*) DNA polymerase, *Bacillus sterothermophilus* (*Bst*) DNA polymerase, *Sulfolobus acidocaldarius* (*Sac*) DNA polymerase, *Thermoplasma acidophilum* (*Tac*) DNA polymerase, *Thermus flavus* (*Tfl/Tub*) DNA polymerase, *Thermus ruber* (*Tru*) DNA polymerase, *Thermus brockianus* (DYNAZYME®) DNA polymerase, *Methanobacterium thermoautotrophicum* (*Mth*) DNA polymerase, and mutants, variants and derivatives thereof.

It will be understood by one of ordinary skill in the relevant arts that other suitable modifications and adaptations to the methods and applications described herein are readily apparent and may be made without departing from the scope of the invention or any embodiment thereof. Having now described the present invention in detail, the same will be more clearly understood by reference to the following examples, which are included herewith for purposes of illustration only and are not intended to be limiting of the invention.

### Examples

The present recombinational cloning method accomplishes the exchange of nucleic acid segments to render something useful to the user, such as a change of cloning vectors. These segments must be flanked on both sides by recombination signals that are in the proper orientation with respect to one another. In the examples below the two parental nucleic acid molecules (*e.g.*, plasmids) are called the Insert Donor and the Vector Donor. The Insert Donor contains a segment that will become joined to a new vector contributed by the Vector Donor. The recombination intermediate(s) that contain(s) both starting

molecules is called the Cointegrate(s). The second recombination event produces two daughter molecules, called the Product (the desired new clone) and the Byproduct.

## 5 **Buffers**

Various known buffers can be used in the reactions of the present invention. For restriction enzymes, it is advisable to use the buffers recommended by the manufacturer. Alternative buffers can be readily found in the literature or can be devised by those of ordinary skill in the art.

10 **Examples 1-3.** One exemplary buffer for lambda integrase is comprised of 50 mM Tris-HCl, at pH 7.5-7.8, 70 mM KCl, 5 mM spermidine, 0.5 mM EDTA, and 0.25 mg/ml bovine serum albumin, and optionally, 10% glycerol.

One preferred buffer for P1 Cre recombinase is comprised of 50 mM Tris-HCl at pH 7.5, 33 mM NaCl, 5 mM spermidine, and 0.5 mg/ml bovine serum albumin.

15 The buffer for other site-specific recombinases which are similar to lambda Int and P1 Cre are either known in the art or can be determined empirically by the skilled artisans, particularly in light of the above-described buffers.

## 20 **Example 1: Recombinational Cloning Using Cre and Cre & Int**

Two pairs of plasmids were constructed to do the *in vitro* recombinational cloning method in two different ways. One pair, pEYC705 and pEYC726 (Figure 2A), was constructed with *loxP* and *att* sites, to be used with Cre and  $\lambda$  integrase. The other pair, pEYC602 and pEYC629 (Figure 3A), contained the *loxP* (wild type) site for Cre, and a second mutant *lox* site, *loxP 511*, which differs from *loxP* in one base (out of 34 total). The minimum requirement for recombinational cloning of the present invention is two recombination sites in each plasmid, in general X and Y, and X' and Y'. Recombinational cloning takes place if either or both types of site can recombine to form a Cointegrate (e.g. X and X'), and if either or both can recombine to excise the Product and Byproduct plasmids

25

30

from the Cointegrate (*e.g.*  $Y$  and  $Y'$ ). It is important that the recombination sites on the same plasmid do not recombine. It was found that the present recombinational cloning could be done with Cre alone.

### Cre-Only

Two plasmids were constructed to demonstrate this conception (*see* Figure 3A). pEZC629 was the Vector Donor plasmid. It contained a constitutive drug marker (chloramphenicol resistance), an origin of replication, *loxP* and *loxP 511* sites, a conditional drug marker (kanamycin resistance whose expression is controlled by the operator/promoter of the tetracycline resistance operon of transposon *Tn10*), and a constitutively expressed gene for the tet repressor protein, *tetR*. *E. coli* cells containing pEZC629 were resistant to chloramphenicol at 30  $\mu\text{g/ml}$ , but sensitive to kanamycin at 100  $\mu\text{g/ml}$ . pEZC602 was the Insert Donor plasmid, which contained a different drug marker (ampicillin resistance), an origin, and *loxP* and *loxP 511* sites flanking a multiple cloning site.

This experiment was comprised of two parts as follows:

**Part I:** About 75 ng each of pEZC602 and pEZC629 were mixed in a total volume of 30  $\mu\text{l}$  of Cre buffer (50 mM Tris-HCl pH 7.5, 33 mM NaCl, 5 mM spermidine-HCl, 500  $\mu\text{g/ml}$  bovine serum albumin). Two 10  $\mu\text{l}$  aliquots were transferred to new tubes. One tube received 0.5  $\mu\text{l}$  of Cre protein (approx. 4 units per  $\mu\text{l}$ ; partially purified according to Abremski and Hoess, *J. Biol. Chem.* 259:1509 (1984)). Both tubes were incubated at 37°C for 30 minutes, then 70°C for 10 minutes. Aliquots of each reaction were diluted and transformed into DH5 $\alpha$ . Following expression, aliquots were plated on 30  $\mu\text{g/ml}$  chloramphenicol; 100  $\mu\text{g/ml}$  ampicillin plus 200  $\mu\text{g/ml}$  methicillin; or 100  $\mu\text{g/ml}$  kanamycin.

**Results:** *See* Table 1. The reaction without Cre gave  $1.1 \times 10^6$  ampicillin resistant colonies (from the Insert Donor plasmid pEZC602);  $7.8 \times 10^5$  chloramphenicol resistant colonies (from the Vector Donor plasmid pEZC629); and 140 kanamycin resistant colonies (background). The reaction with added Cre gave  $7.5 \times 10^5$  ampicillin resistant colonies (from the Insert Donor plasmid pEZC602);  $6.1 \times 10^5$  chloramphenicol resistant colonies (from the Vector Donor plasmid pEZC629);

and 760 kanamycin resistant colonies (mixture of background colonies and colonies from the recombinational cloning Product plasmid). **Analysis:** Because the number of colonies on the kanamycin plates was much higher in the presence of Cre, many or most of them were predicted to contain the desired Product plasmid.

Table 1				
Enzyme	Ampicillin	Chloramphenicol	Kanamycin	Efficiency
None	$1.1 \times 10^6$	$7.8 \times 10^5$	140	$140/7.8 \times 10^5 = 0.02\%$
Cre	$7.5 \times 10^5$	$6.1 \times 10^5$	760	$760/6.1 \times 10^5 = 0.12\%$

**Part II:** Twenty four colonies from the "+ Cre" kanamycin plates were picked and inoculated into medium containing 100 µg/ml kanamycin. Minipreps were done, and the miniprep DNAs, uncut or cut with *SmaI* or *HindIII*, were electrophoresed. **Results:** 19 of the 24 minipreps showed supercoiled plasmid of the size predicted for the Product plasmid. All 19 showed the predicted *SmaI* and *HindIII* restriction fragments. **Analysis:** The Cre only scheme was demonstrated. Specifically, it was determined to have yielded about 70% (19 of 24) Product clones. The efficiency was about 0.1% (760 kanamycin resistant clones resulted from  $6.1 \times 10^5$  chloramphenicol resistant colonies).

### **Cre Plus Integrase**

The plasmids used to demonstrate this method are exactly analogous to those used above, except that pEYC726, the Vector Donor plasmid, contained an *attP* site in place of *loxP 511*, and pEYC705, the Insert Donor plasmid, contained an *attB* site in place of *loxP 511* (Figure 2A).

This experiment was comprised of three parts as follows:

**Part I:** About 500 ng of pEYC705 (the Insert Donor plasmid) was cut with *ScaI*, which linearized the plasmid within the ampicillin resistance gene. (This was done because the λ integrase reaction has been historically done with the *attB* plasmid in a linear state (H. Nash, personal communication). However, it was

found later that the integrase reaction proceeds well with both plasmids supercoiled.) Then, the linear plasmid was ethanol precipitated and dissolved in 20  $\mu$ l of  $\lambda$  integrase buffer (50 mM Tris-HCl, about pH 7.8, 70 mM KCl, 5 mM spermidine-HCl, 0.5 mM EDTA, 250  $\mu$ g/ml bovine serum albumin). Also, about 500 ng of the Vector Donor plasmid pEZC726 was ethanol precipitated and dissolved in 20  $\mu$ l  $\lambda$  integrase buffer. Just before use,  $\lambda$  integrase (2  $\mu$ l, 393  $\mu$ g/ml) was thawed and diluted by adding 18  $\mu$ l cold  $\lambda$  integrase buffer. One  $\mu$ l IHF (integration host factor, 2.4 mg/ml, an accessory protein) was diluted into 150  $\mu$ l cold  $\lambda$  integrase buffer. Aliquots (2  $\mu$ l) of each DNA were mixed with  $\lambda$  integrase buffer, with or without 1  $\mu$ l each  $\lambda$  integrase and IHF, in a total of 10  $\mu$ l. The mixture was incubated at 25°C for 45 minutes, then at 70°C for 10 minutes. Half of each reaction was applied to an agarose gel. **Results:** In the presence of integrase and IHF, about 5% of the total DNA was converted to a linear Cointegrate form. **Analysis:** Activity of integrase and IHF was confirmed.

**Part II:** Three microliters of each reaction (*i.e.*, with or without integrase and IHF) were diluted into 27  $\mu$ l of Cre buffer (above), then each reaction was split into two 10  $\mu$ l aliquots (four altogether). To two of these reactions, 0.5  $\mu$ l of Cre protein (above) were added, and all reactions were incubated at 37°C for 30 minutes, then at 70°C for 10 minutes. TE buffer (90  $\mu$ l; TE: 10 mM Tris-HCl, pH 7.5, 1 mM EDTA) was added to each reaction, and 1  $\mu$ l each was transformed into *E. coli* DH5 $\alpha$ . The transformation mixtures were plated on 100  $\mu$ g/ml ampicillin plus 200  $\mu$ g/ml methicillin; 30  $\mu$ g/ml chloramphenicol; or 100  $\mu$ g/ml kanamycin. **Results:** See Table 2.

Table 2				
Enzyme	Ampicillin	Chloramphenicol	Kanamycin	Efficiency
None	990	20000	4	$4 / 2 \times 10^4 = 0.02\%$
Cre only	280	3640	0	0
Integrase * only	1040	27000	9	$9 / 2.7 \times 10^4 = 0.03\%$
Integrase * + Cre	110	1110	76	$76 / 1.1 \times 10^3 = 6.9\%$

\* Integrase reactions also contained IHF.

**Analysis:** The Cre protein impaired transformation. When adjusted for this effect, the number of kanamycin resistant colonies, compared to the control reactions, increased more than 100 fold when both Cre and Integrase were used. This suggests a specificity of greater than 99%.

**Part III:** 38 colonies were picked from the Integrase plus Cre plates, miniprep DNAs were made and cut with *HindIII* to give diagnostic mapping information. **Result:** All 38 had precisely the expected fragment sizes. **Analysis:** The Cre plus  $\lambda$  integrase method was observed to have much higher specificity than Cre-alone. **Conclusion:** The Cre plus  $\lambda$  integrase method was demonstrated. Efficiency and specificity were much higher than for Cre only.

**Example 2: Using in vitro Recombinational Cloning to Subclone the Chloramphenicol Acetyl Transferase Gene into a Vector for Expression in Eukaryotic Cells (Figure 4A)**

An Insert Donor plasmid, pEYC843, was constructed, comprising the chloramphenicol acetyl transferase gene of *E. coli*, cloned between *loxP* and *attB* sites such that the *loxP* site was positioned at the 5'-end of the gene (Figure 4B). A Vector Donor plasmid, pEYC1003, was constructed, which contained the cytomegalovirus eukaryotic promoter apposed to a *loxP* site (Figure 4C). One microliter aliquots of each supercoiled plasmid (about 50 ng crude miniprep DNA) were combined in a ten microliter reaction containing equal parts of lambda

integrase buffer (50 mM Tris-HCl, pH 7.8, 70 mM KCl, 5 mM spermidine, 0.5 mM EDTA, 0.25 mg/ml bovine serum albumin) and Cre recombinase buffer (50 mM Tris-HCl, pH 7.5, 33 mM NaCl, 5 mM spermidine, 0.5 mg/ml bovine serum albumin), two units of Cre recombinase, 16 ng integration host factor, and 32 ng lambda integrase. After incubation at 30°C for 30 minutes and 75°C for 10 minutes, one microliter was transformed into competent *E. coli* strain DH5α (Life Technologies, Inc.). Aliquots of transformations were spread on agar plates containing 200 μg/ml kanamycin and incubated at 37°C overnight. An otherwise identical control reaction contained the Vector Donor plasmid only. The plate receiving 10% of the control reaction transformation gave one colony; the plate receiving 10% of the recombinational cloning reaction gave 144 colonies. These numbers suggested that greater than 99% of the recombinational cloning colonies contained the desired product plasmid. Miniprep DNA made from six recombinational cloning colonies gave the predicted size plasmid (5026 base pairs), CMVProd. Restriction digestion with *Nco*I gave the fragments predicted for the chloramphenicol acetyl transferase cloned downstream of the CMV promoter for all six plasmids.

**Example 3: Subcloned DNA Segments Flanked by attB Sites Without Stop Codons**

**Part I: Background**

The above examples are suitable for transcriptional fusions, in which transcription crosses recombination sites. However, both attR and loxP sites contain multiple stop codons on both strands, so translational fusions can be difficult, where the coding sequence must cross the recombination sites, (only one reading frame is available on each strand of loxP sites) or impossible (in attR or attL).

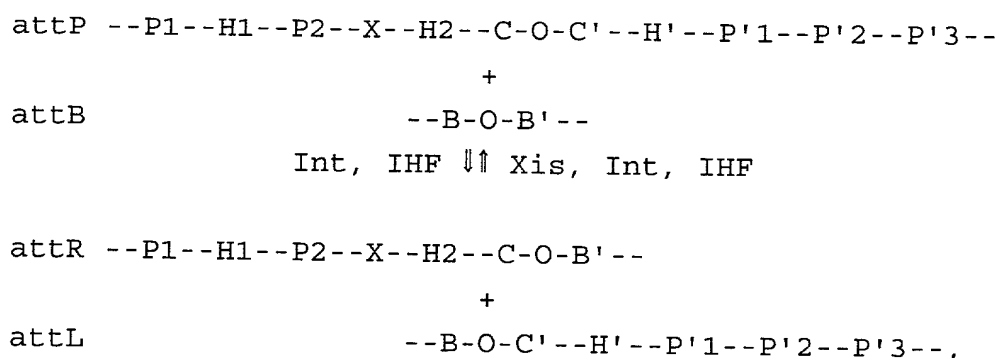
A principal reason for subcloning is to fuse protein domains. For example, fusion of the glutathione S-transferase (GST) domain to a protein of interest allows the fusion protein to be purified by affinity chromatography on glutathione agarose (Pharmacia, Inc., 1995 catalog). If the protein of interest is fused to runs



of consecutive histidines (for example His6), the fusion protein can be purified by affinity chromatography on chelating resins containing metal ions (Qiagen, Inc.). It is often desirable to compare amino terminal and carboxy terminal fusions for activity, solubility, stability, and the like.

The attB sites of the bacteriophage  $\lambda$  integration system were examined as an alternative to loxP sites, because they are small (25 bp) and have some sequence flexibility (Nash, H.A. *et al.*, *Proc. Natl. Acad. Sci. USA* 84:4049-4053 (1987). It was not previously suggested that multiple mutations to remove all stop codes would result in useful recombination sites for recombinational subcloning.

Using standard nomenclature for site specific recombination in lambda bacteriophage (Weisber, in *Lambda III*, Hendrix, *et al.*, eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989)), the nucleotide regions that participate in the recombination reaction in an *E. coli* host cell are represented as follows:



where: O represents the 15 bp core DNA sequence found in both the phage and *E. coli* genomes; B and B' represent approximately 5 bases adjacent to the core in the *E. coli* genome; and P1, H1, P2, X, H2, C, C', H', P'1, P'2, and P'3 represent known DNA sequences encoding protein binding domains in the bacteriophage  $\lambda$  genome.

The reaction is reversible in the presence of the protein Xis (excisionase); recombination between attL and attR precisely excise the  $\lambda$  genome from its integrated state, regenerating the circular  $\lambda$  genome containing attP and the linear *E. coli* genome containing attB.

## Part II: Construction and Testing of Plasmids Containing Mutant att Sites

Mutant attL and attR sites were constructed. Importantly, Landy *et al.* (*Ann. Rev. Biochem.* 58:913 (1989)) observed that deletion of the P1 and H1 domains of attP facilitated the excision reaction and eliminated the integration reaction, thereby making the excision reaction irreversible. Therefore, as mutations were introduced in attR, the P1 and H1 domains were also deleted. attR' sites in the present example lack the P1 and H1 regions and have the NdeI site removed (base 27630 changed from C to G), and contain sequences corresponding to bacteriophage  $\lambda$  coordinates 27619-27738 (GenBank release 92.0, bg:LAMCG, "Complete Sequence of Bacteriophage Lambda").

The sequence of attB produced by recombination of wild type attL and attR sites is:

	B	O	B'	
attBwt:	5' AGCCT	GCTTTTTTATAC <u>TAA</u>	CT <u>TGA</u> 3'	(SEQ. ID NO:60)
	3' TCGGA	CGAAAAAA <u>TATGATT</u>	GAAC 5'	(SEQ. ID NO:44)

The stop codons are italicized and underlined. **Note** that sequences of attL, attR, and attP can be derived from the attB sequence and the boundaries of bacteriophage  $\lambda$  contained within attL and attR (coordinates 27619 to 27818).

When mutant attR1 (attR') and attL1 sites were recombined the sequence attB1 was produced (mutations in **bold, large font**):

	B	O	B'	
attB1:	5' AGCCT	GCTTTTTTT <b>G</b> TAC <b>A</b> AA	CTTG <b>T</b> <sub>3</sub> '	(SEQ. ID NO:6)
	3' TCGGA	CGAAAAAA <b>C</b> ATG <b>T</b> TT	GAAC <b>A</b> <sub>5</sub> '	(SEQ. ID NO:45)

Note that the four stop codons are gone.

5



15

20.

Table 3

Vector donor att site	Gene donor att site	# of kanamycin resistant colonies*
attR ' wt (pEZC1301)	None	1 (background)
"	attLwt (pEZC1313)	147
"	attL1 (pEZC1317)	47
"	attL2 (pEZC1321)	0
attR ' 1 (pEZC1305)	None	1 (background)
"	attLwt (pEZC1313)	4
"	attL1 (pEZC1317)	128
"	attL2 (pEZC1321)	0
attR ' 2 (pEZC1309)	None	0 (background)
"	attLwt (pEZC1313)	0
"	attL1 (pEZC1317)	0
"	attL2 (pEZC1321)	209

(\*1% of each transformation was spread on a kanamycin plate.)

The above data show that whereas the wild type att and att1 sites recombine to a small extent, the att1 and att2 sites do not recombine detectably with each other.

**Part III.** Recombination was demonstrated when the core region of both attB sites flanking the DNA segment of interest did not contain stop codons. The physical state of the participating plasmids was discovered to influence recombination efficiency.

The appropriate att sites were moved into pEZC705 and pEZC726 to make the plasmids pEZC1405 (Figure 5G) (attR ' 1 and attR ' 2) and pEZC1502 (Figure 5H) (attL1 and attL2). The desired DNA segment in this experiment was a copy of the chloramphenicol resistance gene cloned between the two attL sites of pEZC1502. Pairs of plasmids were recombined *in vitro* using Int, Xis, and IHF (no Cre because no loxP sites were present). 100 ng of each plasmid were incubated in 10 µl reactions of 50 mM Tris HCl pH about 7.8, 16.5 mM NaCl, 35 mM KCl, 5 mM spermidine, 0.25 mM EDTA, 0.375 mg/ml BSA, 3% glycerol that contained 8.1 ng IHF, 43 ng Int, 4.3 ng Xis, and 2 units Cre. Reactions were incubated at 25°C for 45 min., 65°C for 10 min, and 1 µl aliquots were transformed into DH5α cells, and spread on kanamycin plates. The yield of

desired kanamycin resistant colonies was determined when both parental plasmids were circular, or when one plasmid was circular and the other linear as presented in Table 4:

Table 4

Vector donor <sup>1</sup>	Insert donor <sup>1</sup>	Kanamycin resistant colonies <sup>2</sup>
Circular pEYC1405	None	30
Circular pEYC1405	Circular pEYC1502	2680
Linear pEYC1405	None	90
Linear pEYC1405	Circular pEYC1502	172000
Circular pEYC1405	Linear pEYC1502	73000

<sup>1</sup> DNAs were purified with Qiagen columns, concentrations determined by A260, and linearized with XbaI (pEYC1405) or AlwNI (pEYC1502). Each reaction contained 100 ng of the indicated DNA. All reactions (10 µl total) contained 3 µl of enzyme mix (Xis, Int, and IHF). After incubation (45 minutes at 25°, 10 minutes at 65°), one µl was used to transform *E. coli* DH5α cells.

<sup>2</sup> Number of colonies expected if the entire transformation reaction (1 ml) had been plated. Either 100 µl or 1 µl of the transformations were actually plated.

**Analysis:** Recombinational cloning using mutant attR and attL sites was confirmed. The desired DNA segment is subcloned between attB sites that do not contain any stop codons in either strand. The enhanced yield of Product DNA (when one parent was linear) was unexpected because of earlier observations that the excision reaction was more efficient when both participating molecules were supercoiled and proteins were limiting (Nunes-Duby *et al.*, *Cell* 50:779-788 (1987).

**Example 4: Demonstration of Recombinational Cloning Without Inverted Repeats**

**Part I: Rationale**

The above Example 3 showed that plasmids containing inverted repeats of the appropriate recombination sites (for example, attL1 and attL2 in plasmid pEYC1502) (Figure 5H) could recombine to give the desired DNA segment flanked by attB sites without stop codons, also in inverted orientation. A concern

was the *in vivo* and *in vitro* influence of the inverted repeats. For example, transcription of a desired DNA segment flanked by attB sites in inverted orientation could yield a single stranded RNA molecule that might form a hairpin structure, thereby inhibiting translation.

Inverted orientation of similar recombination sites can be avoided by placing the sites in direct repeat arrangement att sites. If parental plasmids each have a wild type attL and wild type attR site, in direct repeat the Int, Xis, and IHF proteins will simply remove the DNA segment flanked by those sites in an intramolecular reaction. However, the mutant sites described in the above Example 3 suggested that it might be possible to inhibit the intramolecular reaction while allowing the intermolecular recombination to proceed as desired.

## ***Part II: Structure of Plasmids Without Inverted Repeats for Recombinational Cloning***

The attR2 sequence in plasmid pEYC1405 (Figure 5G) was replaced with attL2, in the opposite orientation, to make pEYC1603 (Figure 6A). The attL2 sequence of pEYC1502 (Figure 5H) was replaced with attR2, in the opposite orientation, to make pEYC1706 (Figure 6B). Each of these plasmids contained mutations in the core region that make intramolecular reactions between att1 and att2 cores very inefficient (see Example 3, above).

Plasmids pEYC1405, pEYC1502, pEYC1603 and pEYC1706 were purified on Qiagen columns (Qiagen, Inc.). Aliquots of plasmids pEYC1405 and pEYC1603 were linearized with XbaI. Aliquots of plasmids pEYC1502 and pEYC1706 were linearized with AlwNI. One hundred ng of plasmids were mixed in buffer (50 mM Tris HCL pH about 7.8, 16.5 mM NaCl, 35 mM KCl, 5 mM spermidine, 0.25 mM EDTA, 0.375 mg/ml BSA, 3% glycerol) containing Int (43.5 ng), Xis (4.3 ng) and IHF (8.1 ng) in a final volume of 10  $\mu$ l. Reactions were incubated for 45 minutes at 25°C, 10 minutes at 65°C, and 1  $\mu$ l was transformed into *E. coli* DH5 $\alpha$ . After expression, aliquots were spread on agar plates containing 200  $\mu$ g/ml kanamycin and incubated at 37°C.

Results, expressed as the number of colonies per 1  $\mu$ l of recombination reaction are presented in Table 5:

**Table 5**

Vector Donor	Gene Donor	Colonies	Predicted % product
Circular 1405	---	100	---
Circular 1405	Circular 1502	3740	3640/3740 = 97%
Linear 1405	---	90	---
Linear 1405	Circular 1502	172,000	171,910/172,000 = 99.9%
Circular 1405	Linear 1502	73,000	72,900/73,000 = 99.9%
Circular 1603	---	80	---
Circular 1603	Circular 1706	410	330/410 = 80%
Linear 1603	---	270	---
Linear 1603	Circular 1706	7000	6730/7000 = 96%
Circular 1603	Linear 1706	10,800	10,530/10,800 = 97%

**Analysis.** In all configurations, i.e., circular or linear, the pEJC1405 x pEJC1502 pair (with att sites in inverted repeat configuration) was more efficient than pEJC1603 x pEJC1706 pair (with att sites mutated to avoid hairpin formation). The pEJC1603 x pEJC1706 pair gave higher backgrounds and lower efficiencies than the pEJC1405 x pEJC1502 pair. While less efficient, 80% or more of the colonies from the pEJC1603 x pEJC1706 reactions were expected to contain the desired plasmid product. Making one partner linear stimulated the reactions in all cases.

### ***Part III: Confirmation of Product Plasmids' Structure***

Six colonies each from the linear pEJC1405 (Figure 5G) x circular pEJC1502 (Figure 5H), circular pEJC1405 x linear pEJC1502, linear pEJC1603 (Figure 6A) x circular pEJC1706 (Figure 6B), and circular pEJC1603 x linear pEJC1706 reactions were picked into rich medium and miniprep DNAs were prepared. Diagnostic cuts with Ssp I gave the predicted restriction fragments for all 24 colonies.

**Analysis.** Recombination reactions between plasmids with mutant attL and attR sites on the same molecules gave the desired plasmid products with a high degree of specificity.

### ***Example 5: Recombinational Cloning with a Toxic Gene***

#### ***Part I: Background***

Restriction enzyme DpnI recognizes the sequence GATC and cuts that sequence only if the A is methylated by the dam methylase. Most commonly used *E. coli* strains are dam<sup>+</sup>. Expression of DpnI in dam<sup>+</sup> strains of *E. coli* is lethal because the chromosome of the cell is chopped into many pieces. However, in dam<sup>-</sup> cells expression of DpnI is innocuous because the chromosome is immune to DpnI cutting.

In the general recombinational cloning scheme, in which the vector donor contains two segments C and D separated by recombination sites, selection for the desired product depends upon selection for the presence of segment D, and the absence of segment C. In the original Example segment D contained a drug resistance gene (Km) that was negatively controlled by a repressor gene found on segment C. When C was present, cells containing D were not resistant to kanamycin because the resistance gene was turned off.

The DpnI gene is an example of a toxic gene that can replace the repressor gene of the above embodiment. If segment C expresses the DpnI gene product, transforming plasmid CD into a dam<sup>+</sup> host kills the cell. If segment D is transferred to a new plasmid, for example by recombinational cloning, then selecting for the drug marker will be successful because the toxic gene is no longer present.

#### ***Part II: Construction of a Vector Donor Using DpnI as a Toxic Gene***

The gene encoding DpnI endonuclease was amplified by PCR using primers 5'CCA CCA CAA ACG CGT CCA TGG AAT TAC ACT TTA ATT TAG3' (SEQ. ID NO: 17) and 5'CCA CCA CAA GTC GAC GCA TGC CGA



CAGCCTTCC AAA TGT3' (SEQ ID NO:18) and a plasmid containing the DpnI gene (derived from plasmids obtained from Sanford A. Lacks, Brookhaven National Laboratory, Upton, New York; also available from American Type Culture Collection as ATCC 67494) as the template.

Additional mutations were introduced into the B and B' regions of attL and attR', respectively, by amplifying existing attL and attR' domains with primers containing the desired base changes. Recombination of the mutant attL3 (made with oligo Xis115) and attR'3 (attR', made with oligo Xis112) yielded attB3 with the following sequence (differences from attB1 in bold):

B	O	B'	
ACCCA	GCTTTCTTGTACAAA	GTGGT	(SEQ ID NO: 8)
TGGGT	CGAAAGAACATGTTT	CACCA	(SEQ ID NO: 47)

The attL3 sequence was cloned in place of attL2 of an existing Gene Donor plasmid to give the plasmid pEYC2901 (Figure 7A). The attR'3 sequence was cloned in place of attR'2 in an existing Vector Donor plasmid to give plasmid pEYC2913 (Figure 7B). The DpnI gene was cloned into plasmid pEYC2913 to replace the tet repressor gene. The resulting Vector Donor plasmid was named pEYC3101 (Figure 7C). When pEYC3101 was transformed into the dam<sup>-</sup> strain SCS110 (Stratagene), hundreds of colonies resulted. When the same plasmid was transformed into the dam<sup>+</sup> strain DH5 $\alpha$ , only one colony was produced, even though the DH5 $\alpha$  cells were about 20 fold more competent than the SCS110 cells. When a related plasmid that did not contain the DpnI gene was transformed into the same two cell lines, 28 colonies were produced from the SCS110 cells, while 448 colonies resulted from the DH5 $\alpha$  cells. This is evidence that the Dpn I gene is being expressed on plasmid pEYC3101 (Figure 7C), and that it is killing the dam<sup>+</sup> DH5 $\alpha$  cells but not the dam<sup>-</sup> SCS110 cells.

### Part III: Demonstration of Recombinational Cloning Using DpnI Selection

A pair of plasmids was used to demonstrate recombinational cloning with selection for Product dependent upon the toxic gene DpnI. Plasmid pEJC3101 (Figure 7C) was linearized with MluI and reacted with circular plasmid pEJC2901 (Figure 7A). A second pair of plasmids using selection based on control of drug resistance by a repressor gene was used as a control: plasmid pEJC1802 (Figure 7D) was linearized with XbaI and reacted with circular plasmid pEJC1502 (Figure 5H). Eight microliter reactions containing buffer ( 50 mM Tris HCl pH about 7.8, 16.5 mM NaCl, 35 mM KCl, 5 mM spermidine, 0.375 mg/ml BSA, 0.25 mM EDTA, 2.5% glycerol) and proteins Xis (2.9 ng), Int (29 ng), and IHF (5.4 ng) were incubated for 45 minutes at 25°C, then 10 minutes at 75°C, and 1  $\mu$ l aliquots were transformed into DH5 $\alpha$  (i.e., dam+) competent cells, as presented in Table 6.

Table 6

Reaction #	Vector donor	Basis of selection	Insert donor	Colonies
1	pEJC3101/Mlu	Dpn I toxicity	---	3
2	pEJC3101/Mlu	Dpn I toxicity	Circular pEJC2901	4000
3	pEJC1802/Xba	Tet repressor	---	0
4	pEJC1802/Xba	Tet repressor	Circular pEJC1502	12100

Miniprep DNAs were prepared from four colonies from reaction #2, and cut with restriction enzyme Ssp I. All gave the predicted fragments.

**Analysis:** Subcloning using selection with a toxic gene was demonstrated. Plasmids of the predicted structure were produced.

**Example 6: Cloning of Genes with Uracil DNA Glycosylase and Subcloning of the Genes with Recombinational Cloning to Make Fusion Proteins**

**Part I: Converting an Existing Expression Vector to a Vector Donor for Recombinational Cloning**

A cassette useful for converting existing vectors into functional Vector Donors was made as follows. Plasmid pEYC3101 (Figure 7C) was digested with *Apal* and *KpnI*, treated with T4 DNA polymerase and dNTPs to render the ends blunt, further digested with *SmaI*, *HpaI*, and *AlwNI* to render the undesirable DNA fragments small, and the 2.6 kb cassette containing the attR' 1 - Cm<sup>R</sup> - Dpn I - attR' -3 domains was gel purified. The concentration of the purified cassette was estimated to be about 75 ng DNA/ $\mu$ l.

Plasmid pGEX-2TK (Figure 8A) (Pharmacia) allows fusions between the protein glutathione S transferase and any second coding sequence that can be inserted in frame in its multiple cloning site. pGEX-2TK DNA was digested with *SmaI* and treated with alkaline phosphatase. About 75 ng of the above purified DNA cassette was ligated with about 100 ng of the pGEX-2TK vector for 2.5 hours in a 5  $\mu$ l ligation, then 1  $\mu$ l was transformed into competent *E. coli* BRL 3056 cells (a *dam*<sup>-</sup> derivative of DH10B; *dam*<sup>-</sup> strains commercially available include DM1 from Life Technologies, Inc., and SCS 110 from Stratagene). Aliquots of the transformation mixture were plated on LB agar containing 100  $\mu$ g/ml ampicillin (resistance gene present on pGEX-2TK) and 30  $\mu$ g/ml chloramphenicol (resistance gene present on the DNA cassette). Colonies were picked and miniprep DNAs were made. The orientation of the cassette in pGEX-2TK was determined by diagnostic cuts with *EcoRI*. A plasmid with the desired orientation was named pEYC3501 (Figure 8B).

**Part II: Cloning Reporter Genes Into an Recombinational Cloning Gene Donor Plasmid in Three Reading Frames**

Uracil DNA glycosylase (UDG) cloning is a method for cloning PCR amplification products into cloning vectors (U.S. patent No. 5,334,515, entirely incorporated herein by reference). Briefly, PCR amplification of the desired DNA

segment is performed with primers that contain uracil bases in place of thymidine bases in their 5' ends. When such PCR products are incubated with the enzyme UDG, the uracil bases are specifically removed. The loss of these bases weakens base pairing in the ends of the PCR product DNA, and when incubated at a suitable temperature (e.g., 37°C), the ends of such products are largely single stranded. If such incubations are done in the presence of linear cloning vectors containing protruding 3' tails that are complementary to the 3' ends of the PCR products, base pairing efficiently anneals the PCR products to the cloning vector. When the annealed product is introduced into *E. coli* cells by transformation, *in vivo* processes efficiently convert it into a recombinant plasmid.

UDG cloning vectors that enable cloning of any PCR product in all three reading frames were prepared from pEZC3201 (Figure 8K) as follows. Eight oligonucleotides were obtained from Life Technologies, Inc. (all written 5' → 3': **rf1 top** (GGCC GAT TAC GAT ATC CCA ACG ACC GAA AAC CTG TAT TTT CAG GGT) (SEQ. ID NO:19), **rf1 bottom** (CAG GTT TTC GGT CGT TGG GAT ATC GTA ATC)(SEQ. ID NO:20), **rf2 top** (GGCCA GAT TAC GAT ATC CCA ACG ACC GAA AAC CTG TAT TTT CAG GGT)(SEQ. ID NO:21), **rf2 bottom** (CAG GTT TTC GGT CGT TGG GAT ATC GTA ATC T)(SEQ. ID NO:22), **rf3 top** (GGCCAA GAT TAC GAT ATC CCA ACG ACC GAA AAC CTG TAT TTT CAG GGT)(SEQ. ID NO:23), **rf3 bottom** (CAG GTT TTC GGT CGT TGG GAT ATC GTA ATC TT)(SEQ. ID NO:24), **carboxy top** (ACC GTT TAC GTG GAC)(SEQ. ID NO:25) and **carboxy bottom** (TCGA GTC CAC GTA AAC GGT TCC CAC TTA TTA)(SEQ. ID NO:26). The rf1, 2, and 3 top strands and the carboxy bottom strand were phosphorylated on their 5' ends with T4 polynucleotide kinase, and then the complementary strands of each pair were hybridized. Plasmid pEZC3201 (Figure 8K) was cut with NotI and SalI, and aliquots of cut plasmid were mixed with the carboxy-oligo duplex (Sal I end) and either the rf1, rf2, or rf3 duplexes (NotI ends) (10 µg cut plasmid (about 5 pmol) mixed with 250 pmol carboxy oligo duplex, split into three 20 µl volumes, added 5 µl (250 pmol) of rf1, rf2, or rf3 duplex and 2 µl = 2 units T4 DNA ligase to each reaction). After 90 minutes

of ligation at room temperature, each reaction was applied to a preparative agarose gel and the 2.1 kb vector bands were eluted and dissolved in 50 µl of TE.

### ***Part III: PCR of CAT and phoA Genes***

5           Primers were obtained from Life Technologies, Inc., to amplify the chloramphenicol acetyl transferase (CAT) gene from plasmid pACYC184, and phoA, the alkaline phosphatase gene from *E. coli*. The primers had 12-base 5' extensions containing uracil bases, so that treatment of PCR products with uracil DNA glycosylase (UDG) would weaken base pairing at each end of the DNAs and allow the 3' strands to anneal with the protruding 3' ends of the rf1, 2, and 3 vectors described above. The sequences of the primers (all written 5' → 3') were: **CAT left**, UAU UUU CAG GGU ATG GAG AAA AAA ATC ACT GGA TAT ACC (SEQ. ID NO:27); **CAT right**, UCC CAC UUA UUA CGC CCC GCC CTG CCA CTC ATC (SEQ. ID NO:28); **phoA left**, UAU UUU CAG GGU ATG CCT GTT CTG GAA AAC CGG (SEQ. ID NO:29); and **phoA right**, UCC CAC UUA UUA TTT CAG CCC CAG GGC GGC TTT C (SEQ. ID NO:30). The primers were then used for PCR reactions using known method steps (*see, e.g.,* U.S. patent No. 5,334,515, entirely incorporated herein by reference), and the polymerase chain reaction amplification products obtained with these primers comprised the CAT or phoA genes with the initiating ATGs but without any transcriptional signals. In addition, the uracil-containing sequences on the amino termini encoded the cleavage site for TEV protease (Life Technologies, Inc.), and those on the carboxy terminal encoded consecutive TAA nonsense codons.

20           Unpurified PCR products (about 30 ng) were mixed with the gel purified, linear rf1, rf2, or rf3 cloning vectors (about 50 ng) in a 10 µl reaction containing 1X REact 4 buffer (LTI) and 1 unit UDG (LTI). After 30 minutes at 37°C, 1 µl aliquots of each reaction were transformed into competent *E. coli* DH5α cells (LTI) and plated on agar containing 50 µg/ml kanamycin. Colonies were picked and analysis of miniprep DNA showed that the CAT gene had been cloned in reading frame 1 (pEZC3601) (Figure 8C), reading frame 2 (pEZC3609) (Figure 8D) and reading frame 3 (pEZC3617) (Figure 8E), and that the phoA gene

had been cloned in reading frame 1 (pEYC3606) (Figure 8F), reading frame 2 (pEYC3613) (Figure 8G) and reading frame 3 (pEYC3621) (Figure 8H).

**Part IV: Subcloning of CAT or phoA from UDG Cloning Vectors into a GST Fusion Vector**

Plasmids encoding fusions between GST and either CAT or phoA in all three reading frames were constructed by recombinational cloning as follows. Miniprep DNA of GST vector donor pEYC3501(Figure 8B) (derived from Pharmacia plasmid pGEX-2TK as described above) was linearized with ClaI. About 5 ng of vector donor were mixed with about 10 ng each of the appropriate circular gene donor vectors containing CAT or phoA in 8 µl reactions containing buffer and recombination proteins Int, Xis, and IHF (Example 5). After incubation, 1 µl of each reaction was transformed into *E. coli* strain DH5α and plated on ampicillin, as presented in Table 7.

**Table 7**

DNA	Colonies (10% of each transformation)
Linear vector donor (pEYC3501/Cla)	0
Vector donor + CAT rf1	110
Vector donor + CAT rf2	71
Vector donor + CAT rf3	148
Vector donor + phoA rf1	121
Vector donor + phoA rf2	128
Vector donor + phoA rf3	31

**Part V: Expression of Fusion Proteins**

Two colonies from each transformation were picked into 2 ml of rich medium (CircleGrow, Bio101 Inc.) in 17 × 100 mm plastic tubes (Falcon 2059, Becton Dickinson) containing 100 µg/ml ampicillin and shaken vigorously for about 4 hours at 37°C, at which time the cultures were visibly turbid. One ml of each culture was transferred to a new tube containing 10 µl of 10% (w/v) IPTG

to induce expression of GST. After 2 hours additional incubation, all cultures had about the same turbidity; the A600 of one culture was 1.5. Cells from 0.35 ml each culture were harvested and treated with sample buffer (containing SDS and  $\beta$ -mercaptoethanol) and aliquots equivalent to about 0.15 A600 units of cells were applied to a Novex 4-20% gradient polyacrylamide gel. Following electrophoresis the gel was stained with Coomassie blue.

**Results:** Enhanced expression of single protein bands was seen for all 12 cultures. The observed sizes of these proteins correlated well with the sizes predicted for GST being fused (through attB recombination sites without stop codons) to CAT (Figure 8I) or phoA (Figure 8J) in three reading frames: CAT rf1 = 269 amino acids; CAT rf2 = 303 amino acids; CAT rf3 = 478 amino acids; phoA rf1 = 282 amino acids; phoA rf2 = 280 amino acids; and phoA rf3 = 705 amino acids.

**Analysis:** Both CAT and phoA genes were subcloned into a GST fusion vector in all three reading frames, and expression of the six fusion proteins was demonstrated.

#### **Example 7: Reverse Recombination and Subcloning by Recombination**

Two plasmids were constructed to demonstrate reverse recombination according to the present invention. The vector pEZC5601 (Figure 10A), containing attB recombination sites and termed the *attB* parent plasmid (this vector may correspond to the Product DNA), further contained an ampicillin resistance gene, an origin of replication, an attB2 site, a tetracycline resistance gene, and an attB0 site, as described above. Plasmid pEZC6701 (Figure 10B), containing attP recombination sites and termed the *attP* parent plasmid (this vector may correspond to the Byproduct DNA or may correspond to a different Vector Donor DNA), also contained a kanamycin resistance gene, an origin of replication, an attP2 site, a gene encoding the toxic protein ccdB, and an attP0 site. Integrase buffer at 10 X concentration comprised 0.25 M Tris HCl pH 7.5, 0.25 M Tris HCl pH 8.0, 0.7 M potassium chloride, 50 mM spermidine HCl, 5mM EDTA, and 2.5 mg /ml BSA. Note that attP0 and attP2 contained the P1

and H1 domains. Integrase (1.5  $\mu$ l of 435 ng/ $\mu$ l) and IHF (1.5  $\mu$ l of 16 ng/ $\mu$ l in 1 X Integrase buffer) were mixed with 13.5  $\mu$ l of 1 X Int buffer to make the recombinase mixture.

Two 8  $\mu$ l reactions were assembled. Reaction A contained 300 ng pEYC6701 plasmid and 2  $\mu$ l of recombinase mixture in 1 X Integrase buffer. Reaction B contained 300 ng pEYC5601, 300 ng pEYC6701, and 2  $\mu$ l of recombinase mixture in 1 X Integrase buffer. Both reactions were incubated at 25°C for 45 minutes, then at 70°C for 5 minutes, and then cooled. TE buffer (792  $\mu$ l of 10 mM Tris HCl pH 7.5, 1 mM EDTA) was added to each reaction, and 1  $\mu$ l of this diluted reaction was transformed into DH5 $\alpha$  UltraMax competent *E. coli* cells (Life Technologies, Inc., Rockville, MD). After 1 hour of expression in non-selective medium, one tenth (100  $\mu$ l) of each transformation was spread onto agar plates containing 100  $\mu$ g/ml kanamycin.

After overnight incubation at 37°C, the plate from reaction A contained 1 colony, while the plate from reaction B contained 392 colonies. Twelve colonies were picked from the reaction B plate into rich liquid medium and grown overnight. Miniprep DNAs prepared from these cultures were run uncut on an agarose gel and all 12 contained a plasmid of about 3.8 kb. Six of the miniprep DNAs were cut with restriction enzyme ClaI and run along with pEYC6701 (the kanamycin resistant parental plasmid) also cut with ClaI. Plasmid pEYC6701 was cut once with ClaI to give a fragment of about 3.8 kb. The six miniprep DNAs cut twice with ClaI to give fragments of about 900 base pairs and about 2900 base pairs.

**Analysis:** Recombination between the attP sites on pEYC6701 and the attB sites on pEYC5601 resulted in the production of two daughter plasmids, the attL product plasmid (Figure 10C) (which may correspond to the Vector Donor DNA or a new Byproduct DNA) that contained the ampicillin resistance and ccdB genes, and the attR product plasmid (Figure 10D) (which may also correspond to the Insert Donor DNA or a new Product DNA) that contained the kanamycin and tetracycline resistance genes. Competent *E. coli* cells that received the attL product plasmid, the attP parent plasmid pEYC6701, or recombination intermediates, were killed by the toxic ccdB gene product. Competent *E. coli* cells



that received the *attB* parent plasmid pEZC5601 were killed by the kanamycin selection. Only competent *E. coli* cells that received the desired *attR* product plasmid, comprising the kanamycin and tetracycline resistance genes, survived to form colonies. The success of the selection strategy was indicated by the large number of colonies from the reaction that contained both parental plasmids, compared to the reaction that contained only one parental plasmid. The reaction mechanism predicted that the desired product plasmid would contain two *Cla*I restriction sites, one in the kanamycin resistance gene from the pEZC6701 *attP* parent plasmid and one in the tetracycline resistance gene from the pEZC5601 *attB* parent plasmid. The presence of the two sites and the sizes of the fragments resulting from the *Cla*I digestion confirmed the reaction mechanism.

Thus, the present invention relates to reversal of the recombination reaction shown in Figure 1, in which the Product DNA and Byproduct DNA may be combined to produce the Insert Donor DNA and the Vector Donor DNA. Additionally, the invention provides for subcloning recombinations, in which a Product DNA (produced according to Figure 1) may be combined with a new Vector Donor DNA to produce a new Product DNA (in a different Vector background) and a new Byproduct.

#### ***Example 8: Subcloning of Linearized Fragments***

Plasmid pEZC7102 (Figure 11A), the *attP* parent plasmid (which may correspond to the Vector Donor DNA), contained segments *attP*1, origin of replication, kanamycin resistance, *attP*3, chloramphenicol resistance, and the toxic gene *ccdB*, and in the experiment described here was supercoiled. Plasmid pEZC7501 (Figure 11B), the *attB* parent plasmid (which may correspond to the Insert Donor DNA or the Product DNA), contained the GFP gene cloned between *attB* 1 and *attB*3 sites in a vector that comprised the functional domains of pCMVSPORT2.0 (Life Technologies, Inc.). The *attP* sites contained the P1 and H1 domains. Plasmid pEZC7501 was used uncut, or was linearized within the ampicillin resistance gene with *Sca*I, or was cut with *Xba*I and *Sal*I, to yield a

fragment comprising the SalI end, 22bp, the attB1 site, the GFP gene, the attB3 site, and 14 bp to the XbaI end:

SalI end -- 22bp -- attB1 -- GFP -- attB3 -- 14bp -- XbaI end

5

Reactions (8 µl final volume) contained about 40 ng of each DNA, 1 X Int buffer (25 mM Tris HCl pH 7.5, 25 mM Tris HCl pH 8.0, 70 mM KCl, 5 mM spermidine HCl, 0.5 mM EDTA, and 0.25 mg/ml BSA), 12.5% glycerol, 8 ng IHF, and 43 ng lambda integrase. Reactions were incubated at 25 °C for 45 minutes, then at 70°C for 5 minutes, and then cooled. Duplicate 1 µl aliquots of each reaction were transformed into DH5α UltraMax cells and plated in duplicate on kanamycin agar plates.

10

15

The reaction that contained only (supercoiled) pEYC7102 gave an average of 2 colonies (range 1 to 4). The reaction that contained both pEYC7102 and supercoiled pEYC7501 gave an average of 612 colonies (range 482 - 762). The reaction that contained pEYC7102 and linear (ScaI-cut) pEYC7501 gave an average of 360 colonies (range 127-605). The reaction that contained pEYC7102 and the GFP gene on a fragment with attB sites and 22bp and 14 bp beyond the attB sites (pEYC7501 cut with SalI and XbaI) gave an average of 274 colonies (range 243-308).

20

Miniprep DNAs were prepared from 4 colonies from the pEYC7102 x supercoiled pEYC7510 reaction, and from 10 colonies from the pEYC7102 x pEYC7501/SalI + XbaI reaction. All 14 DNAs were run uncut on an agarose gel, and the 10 DNAs from the pEYC7102 x pEYC7501/SalI + XbaI reaction were cut with a mixture of NcoI and PstI and run on an agarose gel. All the uncut plasmids were about 2.8 kb in size. All ten plasmids cut with the mixture of NcoI and PstI gave fragments of about 700 and 2100 bp.

25

The results are presented in Table 8:

30

Table 8

attP Parent	attB Parent	Colonies (average of 4 plates)	Minipreps done	Uncut product plasmid size	Fragment sizes, Nco + Pst digest
sc 7102	--	2	--	--	--
sc 7102	sc 7501	612	4	2.8 kb	--
sc 7102	7501/ScaI	360	--	--	--
sc 7102	7501/SalI + XbaI	274	10	2.8 kb	ca. 2100 bp, 700 bp

**Analysis:** It was expected that the integrative reaction between the attB sites on plasmid pEYC7501 and the attP sites on plasmid pEYC7102 would produce the *attL* product plasmid (Figure 11C) (corresponding to the Insert Donor DNA) containing the GFP segment from pEYC7501, and the kanamycin - origin segment from pEYC7102. The presence of the toxic gene *ccdB* on the *attP* parent plasmid pEYC7102 (corresponding to the Byproduct DNA) was predicted to kill all the cells that received this plasmid. The large increase in the number of colonies when pEYC7501 was present indicated that the desired reaction was occurring, and that the efficiency of the reaction was adequate even if the *attB* parent plasmid (corresponding to the Product DNA) was linear (ScaI cut), or if the attB sites and the GFP gene were present on a fragment that contained little additional sequence beyond the attB sites.

These results show that linear fragments can be suitably subcloned into a different vector by the method of the invention.

### Example 9: Cloning Long PCR Fragments

A PCR product was designed to have an attB0 (wild type) site at one end and a loxP site at the other end. The rationale was that the attP0 x attB0 reaction would go well with the attB0 molecule (the PCR product) linear, (since it involves a normal lambda integration reaction), and that the loxP x loxP excision from the

cointegrate would also be efficient (the unimolecular excision reaction is efficient, the bimolecular integration reaction is inefficient with Cre).

The sequence of the attB-containing PCR primer was 5'-TCC GTT GAA GCC TGC TTT TTT ATA CTA ACT TGA GCG AAG CCT CGG GGT CAG CAT AAG G-3' (SEQ ID NO:31). The sequence of the loxP primer was 5'-CCA ATA ACT TCG TAT AGC ATA CAT TAT ACG AAG TTA TTG CCC CTT GGT GAC ATA CTC G-3' (SEQ ID NO:32). These primers amplify a part of the human myosin heavy chain. Polymerase chain reactions were performed using ELONGASE™ and K562 human DNA as template. Polymerase chain reactions were performed as follows. Reactions (50 microliters) contained 100 ng K562 human DNA (Life Technologies, Inc.), 0.2 μM of each primer, and 0.2 mM of each dNTP, in ELONGASE™ SuperMix (Life Technologies, Inc.). Reactions in thin wall tubes under mineral oil were denatured at 94 °C for 1 minute, then cycled 35 times at 94 °C for 30 seconds, 65 °C for 30 seconds, and 68 °C for 8 minutes 30 seconds. Following thermal cycling, reactions were maintained at 4 °C. The 5.2 kb PCR product (Figure 9A) was gel purified.

Plasmid pEZC1202 (Figure 9B) contained a wild-type attP site, a chloramphenicol resistance gene, a gene encoding the tet repressor, a wild-type loxP site, an origin of replication, and a tet operator/promoter transcribing and controlling the transcription of a kanamycin resistance gene. This plasmid conferred chloramphenicol resistance but not kanamycin resistance, because the tet repressor made by one element of the plasmid kept the kanamycin resistance gene turned off. The pEZC1202 DNA used in this experiment was a miniprep whose concentration was estimated to be about 50 ng per microliter.

About 40 ng of the gel purified 5.2 kb PCR product were included in a 10 μl reaction that contained about 50 ng of supercoiled pEZC1202, 0.2 units of Cre recombinase, 3.6 ng IHF, and 11 ng of Int in 50 mM Tris HCl pH about 7.8, 16 mM NaCl, 35 mM KCl, 0.25 mM EDTA, 0.375 mg/ml bovine serum albumin. A second reaction that did not contain the PCR product was included as a control. After incubating at 27° for 45 min and then 70° for 5 minutes, 1 μl aliquots were transformed into DH5α UltraMax competent *E. coli* cells (Life Technologies, Inc.). One fifth of each expression mix was plated on agar that contained

100 µg/ml kanamycin and the plates were incubated overnight at 37° C. The reaction that contained the PCR product gave 34 colonies, while the reaction that lacked the PCR product gave 31 colonies. After the plates sat at room temperature for four days, 26 additional small colonies were seen on the plate from the positive (+ PCR product) reaction, while only one additional small colony was seen on the plate from the negative (no PCR product) reaction.

Twelve of the 26 small colonies were grown overnight in rich broth (CircleGrow) that contained 25 µg/ml kanamycin, and miniprep DNAs were prepared from these cultures. All twelve miniprep plasmids were about 8 kb in size, which corresponded to the size expected for replacement of the chloramphenicol resistance and tet repressor genes in pEZC1202 with the 5.2 kb PCR product. The predicted recombinant product is shown in Figure 9C. Two of these plasmids were cut with *Ava*I (8 sites predicted) and *Bam*HI (4 sites predicted). All the predicted *Ava*I fragments appeared to be present. One of the *Bam*HI sites predicted in the PCR product (the one closest to the *attB* end) was absent from both minipreps, but the other *Bam*HI fragments were consistent with the expected structure of the cloned 5.2 kb PCR product.

**Analysis:** The replacement of the chloramphenicol resistance and tet repressor genes in pEZC1202 with the 5.2 kb PCR product (part of the human myosin heavy chain) conferred a moderate resistance of the host *E. coli* cells to kanamycin, but this resistance was not sufficient to allow colonies to appear after overnight incubation. Thus, colonies containing the desired recombination product grew on kanamycin plates, but were not seen after overnight incubation, but only after an additional room temperature incubation. Of the 12 *Ava*I and *Bam*HI restriction sites predicted from the nucleotide sequence, 11 were confirmed experimentally. Thus the following three observations support the conclusion that the 5.2 kb PCR product was cloned by recombination: (a) small, slow growing colonies appeared only on the plate from the reaction that contained the PCR product; (b) the miniprep plasmids from these colonies were the expected size; and (c) diagnostic restriction cuts gave the expected fragments (with the one above noted exception).

### Example 10: Cloning of PCR Fragments

Three sets of pairs of PCR primers (Table 9) were designed to amplify an 830 bp sequence within plasmid pEZC7501 (Figure 11B) comprising: attB 1--GFP--attB3, with or without additional nucleotides at the outer ends of the 25 bp attB1 and attB3 recombination sites. (Here "outer" refers to the end of the attB sequence that is not adjacent to the GFP gene sequence.) Primer set A added 17 nucleotides upstream of attB1 and 15 nucleotides downstream of attB3; primer set B added 5 and 8 nucleotides to attB1 and attB3, respectively; and primer set C added no additional nucleotides to either attB recombination sequence.

The primer sequences are provided in Table 9:

Table 9	
upper GFP A	5'-TCA CTA GTC GGC GGC CCA CA (SEQ ID NO:33)
lower GFP A	5'-GAG CGG CCC CCG CGG ACC AC (SEQ ID NO:34)
upper GFP B	5'-GGC CCA CAA GTT TGT ACA AAA (SEQ ID NO:35)
lower GFP B	5'-CCC CGC GGA CCA CTT TGT AC (SEQ ID NO:36)
upper GFP C	5'-ACA AGT TTG TAC AAA AAA GCA (SEQ ID NO:37)
lower GFP C	5'-ACC ACT TTG TAC AAG AAA GCT (SEQ ID NO:38)

### PCR Reactions

Primer sets A and C were used first with the following PCR reactions, in 50 µl, in duplicate. Final concentrations were:

20 mM TrisHCl, pH 8.4

50 mM KCl

0.2mM of all four deoxynucleotide triphosphates (dNTPs)

400 ng/ml pEZC7501 supercoiled DNA template

0.5 µM of each primer

Recombinant Taq DNA polymerase (BRL-GIBCO) 100 U/ml

A duplicate set of the above reactions contained 1 M betaine.

The reactions were first heated for to 94°C for 1', then cycled 25 times at 94°C for 45", 55°C for 30", and 72°C for 1'.

The size of the PCR reaction products was analyzed on a 1% agarose gel in TAE buffer containing 0.5 µg/ml ethidium bromide. All reactions yielded products of the expected size, thus duplicate reactions were pooled. As the corresponding reactions with and without betaine were not significantly different, these also were pooled, giving a final pooled volume for reactions with primer sets A and C of 200 µl each.

Primer set B was then used with identical reactions to those above performed, except that the reaction volumes were increased to 100µl. After duplicate reactions and reactions plus and minus betaine were pooled, the final volume of the reactions with primer set B was 400 µl.

The three pooled primer reaction products were stored at -20°C for 4 weeks.

### ***PCR Product Purification***

Each of the three pooled PCR products was extracted once with an equal volume of a mixture of Tris-buffered phenol, isoamyl alcohol and chloroform. The aqueous supernatant then was extracted twice with an equal volume of isobutanol, and the aqueous layer ethanol precipitated with two volumes of ethanol, 0.1 M sodium acetate, pH 6.2. The ethanol precipitates were recovered by centrifugation at 13,000 rpm for 10' at room temperature, and the supernatant discarded. The dried pellets were dissolved in TE: 100µl for reactions prepared with primer sets A and C; 200 µl for the reactions with primer set B.

To remove PCR primers and extraneous small PCR products, the PCR products were precipitated with polyethylene glycol (PEG) by adding ½ volume of a solution of 30% PEG 8000 (Sigma), 30 mM MgCl<sub>2</sub>, mixing well, and centrifuging at 13,000 rpm for 10', all at room temperature. The supernatant was discarded, and the pellets were dissolved in their previous volume of TE buffer. 1µl aliquots of each of the three PCR products were checked on a 1% agarose gel to quantitate the recovery, which was estimated to be over 90%. The concentration of each PCR product was adjusted using TE to 40 ng/µl.

***Recombination Reaction with the PCR Products of Primer sets A, B, and C***

Five 8 µl reactions were assembled in 1 X Integrase buffer (25 mM Tris HCl pH 7.5, 25 mM Tris HCl pH 8.0, 80 mM KCl, 5 mM spermidine, 0.5 mM EDTA, 0.25 mg/ml BSA) containing: 40 ng of pEZC7102 DNA, 2 µl of recombinase mixture (8 ng/µl IHF, 22 ng/µl Int in 1 X Int Buffer, 50% glycerol) the reactions differed by the addition of either the PCR product of primer set A (reaction A), primer set B (reaction B), or primer set C (reaction C); the addition of no PCR product (reaction D), or the addition of 40 ng of pEZC7501 SC (supercoiled) DNA (reaction E) as a positive control. All reactions were performed in duplicate.

The reactions were incubated for 45' at 25°C, for 10' at 70°C, then held at 0-5°C. 2 µl aliquots of each reaction were transformed into Max Efficiency DH5α, in a 50 µl transformation reaction, and following expression in 50C medium, 1/5 (100 µl) and 4/5 (400 µl) of the reactions were plated on kanamycin-containing (50 µg/ml) LB culture plates. The results of the duplicate reactions are shown in Table 10.

Table 10	
Transfection	No. Colonies
A 100 µl	464, 668
A 400 µl	>1000, >1300
B 100 µl	980, 1292
B 400 µl	>3000, >3000
C 100 µl	2, 8
C 400 µl	13, 20
D 100 µl	0, 0
D 400 µl	0, 0
E 100 µl	56, 70

***Analysis of the colonies obtained***

Miniprep DNA was prepared from 8 colonies of each of the Recombination reactions with primer sets A, B, or C. The supercoiled DNA obtained was analyzed on a 1% agarose gel: all eight of colonies from the



recombination products of primer sets A and B were of the predicted size (2791 bp) for correct recombination between the PCR products (about 824 bp) and the attB1--ori--kan<sup>r</sup>--attB3 sequence donated by pEZC 7102 (1967 bp). Three of the eight reaction products of primer set C were of the predicted size; the other five all were slightly larger than 4 kb.

Further analysis of the reaction products was performed using two different restriction enzymes, *Ava*I and *Pvu*II, each of which cleaves twice (but at different locations) within the predicted recombinant product, once within the PCR product sequence and once within the sequence contributed by pEZC7102. Both of these enzymes should cleave the intact pEZC7102 recombination partner plasmid at two sites, to give fragments easily distinguished from those of the expected recombination products.

The two restriction enzyme digests yielded the expected sizes of fragments (2373 and 430 bp for *Ava*I; 2473 and 330 bp for *Pvu*II) from the colonies generated from the recombination reactions with primer sets A and B, as well as for the three colonies from primer set C that displayed the expected size of supercoiled DNA. For the other five colonies from primer set C that yielded larger SC DNA, however, the *Pvu*II digest revealed fragments of approximate size to those predicted from a digestion of pEZC7102, whereas the *Ava*I digest revealed only a single fragment, approximately the size of linearized pEZC7102 (4161 bp).

### *Analysis*

These results indicate that PCR products generated from templates containing a gene flanked by attB sites can serve as efficient substrates for the reverse recombination reaction. The addition of even short DNA sequences to the ends of the attB1 and attB3 sites or core regions (*e.g.*, 5bp and 8bp, respectively, in primer set B) stimulated this reaction by 100 fold or more. Surprisingly, reverse recombination reactions with PCR products containing additional sequence beyond the attB sites appeared in these reactions to be more efficient recombination partners than the supercoiled positive control plasmid, pEZC7501.

All the recombination products were generated faithfully. A low level of background colonies emerged from the relatively inefficient recombination reactions with primer set C, which lacked additional sequence beyond the 25 bp attB sites. This background appeared to be due to a largely intact pEZC7102 (which encodes kanamycin resistance) lacking an active ccdB death gene, allowing it to survive. Consistent with this interpretation is that one of the two restriction sites for *AvaI* in this plasmid was also altered. One of the *AvaI* sites is present within the ccdB region of pEZC7102. It is likely therefore that the alteration of this site was secondary to mutational inactivation of the ccdB gene.

### ***Example 11: Further Cloning of PCR Fragments***

Two sets of 6 primers for preparing PCR products from the plasmid pBR322 as template were used. One set (Table 11) anneals to sequences flanking the *TetR* gene, including the *TetR* promoter. The other set (Table 12) anneals to sequences flanking the *AmpR* gene, including its promoter. The "tet" and "amp" primers used contain no attB sequences, only sequences inherent to the pBR322 plasmid; the "attB" primers contain, in addition to the pBR322 sequences, the 25 bp of attB 1 or attB3 sequences; the "attB+4" primers contain the pBR322-specific sequences, plus the 25 bp attB 1 or attB3 sequences, each with four Gs at the outer end. (Here "outer" refers to the end of the attB sequence not adjacent to the template-specific primer sequence.)

### ***Preparation of pBR322 template***

To improve the efficiency of the PCR reaction, the supercoiled pBR322 DNA was linearized by incubating 3.5 µg of Supercoiled (SC) pBR322 DNA in a 200 µl reaction with 15 units of the restriction enzyme *NdeI* and final concentration of 50mM Tris-HCl, pH8.0, 10mM MgCl<sub>2</sub>, and 50mM NaCl, for one hour at 37°C.

The digested pBR322 DNA was extracted once with phenol, isoamyl alcohol, and chloroform, extracted twice with isobutanol, and precipitated by adding two volumes of ethanol plus 0.15M sodium acetate. The precipitate was

washed once with 100% ethanol, dried, then dissolved in TE buffer. Recovery of DNA, quantitated on a 1 % agarose gel in TAE buffer, 0.5 µg/ml ethidium bromide, was estimated as greater than 80%.

Table 11

tet Primer	Primer Sequence	SEQ ID NO:
tet-L	AAT TCT CAT GTT TGA CAG CTT ATC	48
tet-R	CGA TGG ATA TGT TCT GCC AAG	49
attB1-tetL	ACAAG TTTGTA CAAAAA AGCA GGCT- AAT TCT CAT GTT TGA CAG CTT ATC	50
attB3-tetR	ACCAC TTTGTA CAAGAA AGCT GGGT- CGA TGG ATA TGT TCT GCC AAG	51
attB1+4-tetL	GGGG ACAAG TTTGTA CAAAAA AGCA- GGCT AAT TCT CAT GTT TGA CAG CTT- ATC	52
attB3+4-tetR	GGGG ACCAC TTTGTA CAAGAA AGCT- GGGT CGA TGG ATA TGT TCT GCC AAG	53

Table 12

amp Primer	Primer Sequence	SEQ ID NO:
amp-L	AAT ACA TTC AAA TAT GTA TCC GC	54
amp-R	TTA CCA ATG CTT AAT CAG TGA G	55
attB1-ampL	ACAAG TTTGTA CAAAAA AGCA GGCT - AAT ACA TTC AAA TAT GTA TCC GC	56
attB3-ampR	ACCAC TTTGTA CAAGAA AGCT GGGT- TTA CCA ATG CTT AAT CAG TGA G	57
attB1+4-ampL	GGGG ACAAG TTTGTA CAAAAA AGCA- GGCT AAT ACA TTC AAA TAT GTA TCC- GC	58
attB3+4-ampR	GGGG ACCAC TTTGTA CAAGAA AGCT- GGGT TTA CCA ATG CTT AAT CAG TGAG	59

### ***PCR amplification of tet and amp gene sequences***

Six PCR reactions were performed, in 100µl, consisting of 20mM Tris-HCl, pH 8.4, 50mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.2mM dNTPs, 2 ng linearized pBR322, 2.5 units of Taq DNA polymerase (GIBCO-BRL), and 0.5 uM of each pair of PCR primers listed in Tables 3 and 4. The reactions were first heated to 94 C for 3'; then subjected to 25 cycles of 94°C for 45 seconds, 55°C for 30 seconds, and 72°C for 1 minute. Based on 1% agarose gel analysis, all the reactions generated products of the expected size, in reasonable yields.

### ***Purification of PCR products***

The products from duplicate reactions were pooled; extracted with an equal volume of phenol, isoamyl alcohol, and chloroform; extracted twice with an equal volume of isobutanol; and precipitated with two volumes of ethanol, as above. The six precipitates were washed once with 100% ethanol, dried and dissolved in 100 µl TE. 1 µl aliquots were taken for gel analysis of the product before PEG precipitation.

To each tube was added 50 µl of 30% PEG 8000, 30 mM MgCl<sub>2</sub>. The solution was mixed well and centrifuged at 13,000 rpm for 10', at room temperature. The supernatant was carefully removed, and the precipitate dissolved in 100 µl TE. Recovery was quantitated on a 1% agarose and estimated to be over 90%. The gel analysis also revealed that nucleic acid products smaller than about 300 nucleotides had been effectively removed by the PEG precipitation step.

### ***Recombination Reactions***

Seven recombination reactions were performed, each in a total volume of 8 µl, containing 1 X integrase buffer, 40 ng pEZC7102 (Figure 11A), and 2 µl recombinase mixture (see above, Example 10). Each of the reactions also contained approximately 40 ng of one of the six above PCR products or, as a positive control, 40 ng of pEZC7501 (Figure 11B). The amp and tet PCR products with attB sites at their termini are shown in Figures 12A and 12B. The

reactions were incubated at 25°C for 45', at 70° C for 10', then held at 0-5°C for 1-2 hours until used to transform *E. coli*.

### *E. coli transformation with recombination reaction products*

1 µl of each of the recombination reactions was transformed into Max Efficiency DH5α in a 50 µl transformation reaction, and following expression in SOC medium, 1/5 (100 µl) and 4/5 (400 µl) of each reaction were plated on culture plates containing 50 µg/ml kanamycin. The plates were incubated overnight and colonies were counted. The number of colonies obtained from each set of duplicate reactions are displayed in Table 13:

Table 13

Recombination Reactions	No. Colonies
tet 100 (100 µl)	6, 10
tet 400 (400 µl)	27, 32
attB-tet 100	9, 6
attB-tet 400	27, 36
attB+4-tet 100	824, 1064
attB+4-tet 400	>2000, >4000
amp 100	7, 13
amp 400	59, 65
attB-amp 100	18, 22
attB-amp 400	66, 66
attB+4-amp 100	3020, 3540
attB+4-amp 400	>5000, >5000
pEYC7501 100	320, 394
pEYC7501 400	1188, 1400

### *Analysis of the colonies obtained*

As a rapid phenotypic screen, 10 of the colonies from the *tet* EYC reactions and 33 of the colonies from the attB+4-tet EYC reactions were streaked onto an LB culture plate containing tetracycline (15 µg/ml). As a control for the potency of the tetracycline, 3 colonies of pUC19-transformed cells, lacking a *TetR* gene, were also streaked onto the plate. All colonies from the attB+4-tet EYC

reactions grew well; colonies from the *tet* EZC reactions grew only very slightly, and the pUC19 colonies grew not at all.

Analogous results were obtained by streaking colonies from the *amp* PCR reactions on culture plates containing ampicillin (100 ug/ml ). All 21 colonies generated from the attB+4-*amp* recombination reactions grew well, whereas only one of 13 colonies from the attB-*amp* reactions grew in the presence of ampicillin. No growth was seen with any of the 15 colonies from the recombination reaction with *amp* PCR products.

To characterize plasmid DNA, eight colonies generated from the six EZC reactions with PCR products were picked into LB broth containing 50 µg/ml kanamycin and grown overnight at 37°C. Miniprep DNA was prepared from 0.9 ml of each culture, and the size of the supercoiled DNA was analyzed on a 1% agarose gel in TAE buffer containing 0.5 µg/ml ethidium bromide. The results are displayed in Table 14. The predicted structures of the recombination products are shown in Figure 12C and 12D.

**Table 14**

Recombination Reactions	DNA	Predicted Size (bp)	Number with Predicted Size
tet	SC (supercoiled)	3386	0/8
attB-tet	SC	3386	1/8
attB+4-tet	SC AvaI+Bam	3386 485, 2901	7/7 3/3
amp	SC	2931	0/8
attB-amp	SC	2931	3/8
attB+4-amp	SC Pst	2931 429, 2502	8/8 3/3

### *Analysis*

These results, based on the amplification of two different gene sequences, tet and amp, within the plasmid pBR322 clearly demonstrate that PCR products generated using primers containing the 25 bp attB1 and attB3 recombination sequence serve as highly efficient substrates for the recombination reaction. Addition of a short sequence to the outside of each 25 bp attB site stimulates the recombination reaction by over 100 fold, as also observed in the experiments of Example 10. Also similar to Example 10, the efficiency of the recombination reactions using linear PCR products with attB sites exceeded the efficiency obtained with the positive control SC DNA plasmid, pEZC7501.

Further, a high percentage of the reaction products are as predicted, since all 33 colonies tested from the attB+4-tet reactions displayed functional tetracycline resistance, and all 21 of the colonies from the attB+4-amp reactions displayed ampicillin resistance. All 16 of the miniprep DNAs, examined from the recombination reactions of either attB+4-tet or attB+4-amp PCR products with pEZC7102, generated supercoiled DNA and restriction digest fragments of the correct sizes.

### *Example 12: Use of Topoisomerase to Stimulate Recombination*

The stimulation of the recombination reaction by making one or the parental plasmids linear was not expected. If the stimulation resulted from relief of some conformation constraint arising during the two recombination reactions (formation of the Cointegrate and resolution to the two daughter molecules), then unwinding of the plasmids with a topoisomerase might also be stimulatory when one or both parental plasmids were circular.

The Insert Donor was pEZC2901 (Figure 7A), and the Vector Donor was pECZ3101 (Figure 7B). A portion of pEZC3101 was linearized with Mlu I. 20 ng of pEZC2901 and/or pECZ3101 were used in each 10  $\mu$ l reaction (29 ng Int, 2.9 ng Xis, 5.4 ng IHF in 50 mM Tris HCl pH about 7.8, 16.5 mM NaCl, 35 mM KCl, 5 mM spermidine, 0.375 mg/ml BSA, 0.25 mM EDTA, 2% glycerol).

Topoisomerase I (from calf Thymus; Life Technologies, Inc.) was diluted from 15 units/ $\mu$ l to the concentrations indicated in Table 15 in 1 X EZC buffer.

**Table 15**

	1	2	3	4	5	6	7	8	9	10
Circular 3101						2	2	2	2	2
Linear 3101	2	2	2	2	2					
Circular 2901		2	2	2	2		2	2	2	2
Recombinase	2	2	2	2	2	2	2	2	2	2
TE	2					2				
Topoisomerase, 1:60			2					2		
Topoisomerase, 1:20				2					2	
Topoisomerase, 1:6					2					2
3 X Buffer	2	2	2	2	2	2	2	2	2	2
1 X Buffer	2	2				2	2			

These reactions were assembled in the following order: buffer; TE; DNAs; Clonase; Topoisomerase. The reactions were incubated at 22° - 28° for 45 minutes, then at 70° for 5 minutes. 1  $\mu$ l aliquots were transformed into UltraMax DH5 $\alpha$  competent *E. coli* (Life Technologies, Inc.). Following expression, aliquots were plated on 100  $\mu$ g/ml kanamycin and incubated at 30° for 48 hours. Results: see Table 16.

**Table 16**

Reaction #	Colonies	Vector Donor	Insert Donor	Recombinase	Topo-isomerase
1	0	linear 3101	--	+	--
2	245	linear 3101	circular 2901	+	--
3	221	linear 3101	circular 2901	+	0.5 units
4	290	linear 3101	circular 2901	+	1.6 units
5	355	linear 3101	circular 2901	+	5 units
6	0	circular 3101		+	--
7	23	circular 3101	circular 2901	+	--
8	209	circular 3101	circular 2901	+	0.5 units
9	119	circular 3101	circular 2901	+	1.6 units
10	195	circular 3101	circular 2901	+	5 units



### *Analysis*

Linearizing the Vector Donor increased the number of colonies about 10 fold (reaction 2 vs. reaction 7). Addition of 0.5 to 5 units of topoisomerase I to reactions containing circular Insert Donor and linear Vector Donor had little or no effect on the number of colonies (reaction 2 compared to reactions 3, 4, and 5; maximum 1.4 fold). In contrast, if both parental plasmids were circular (reaction 7 -10), the addition of topoisomerase stimulated the number of colonies 5 to 9 fold. Thus addition of topoisomerase I to reactions in which both parental plasmids were circular stimulated the recombination reactions nearly as much as linearizing the Vector Donor parent. Topoisomerase I was active when used in combination with the three recombination proteins, in recombination buffer. The addition of topoisomerase I to the recombination reaction relieves the necessity to linearize the Vector Donor to achieve stimulation of the recombination reactions.

Having now fully described the present invention in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious to one of ordinary skill in the art that the same can be performed by modifying or changing the invention within a wide and equivalent range of conditions, formulations and other parameters without affecting the scope of the invention or any specific embodiment thereof, and that such modifications or changes are intended to be encompassed within the scope of the appended claims.

All publications, patents and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains, and are herein incorporated by reference to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference.